Kuwait J. Sci. **45 (4)** pp 26 - 36, 2018

A comprehensive comparison of phase I dose escalation methods

Efehan Ulas[1,2,*], Filiz Karaman[2]

[1]*Dept. of Statistics, Faculty of Science, Cankiri Karatekin University, Cankiri, Turkey*

[2]*Dept. of Statistics, Faculty of Science, Yildiz Technical University, Istanbul, Turkey*

**Corresponding author: ef_ulas@hotmail.com*

Abstract

Phase I clinical trials are fundamental in drug development because they bring proposed designs to initial clinical testing. Recently, several dose finding methods have been developed. However, the comparison of those designs and traditional designs are not intensive. This study compares the most commonly used phase I dose finding methods and determines which one performs better. To do so, two different real life stories are analyzed through simulation studies. It was found that the 3+3 design, the most popular method employed by scientists, produced the worst results. More reliable and applicable results for phase I dose escalation trials can be produced by BMA-CRM, CRM, and BCRM designs.

Keywords: Dose escalation; phase I trials; biomarkers; maximum tolerated dose.

1. Introduction

Clinical trials play a critical role on how new medical approaches work in humans. Moreover, they aim to detect, treat or manage known and unknown diseases or medical conditions. Friedman *et al.* (2010) define a trial as a "prospective study for determining the effect of treatment in humans." Clinical trials are just one of the stages in long and careful processes. Experts have been working for many years to understand the effects of the new treatments and also their side effects. For this reason, clinical trials start with a small group of people to reduce any possible damages. Although there is no certainty that clinical trials will result in favorable treatment, the participating patients provide significant contributions to future treatments. Erroneous decisions can be made if a doctor considers patients with frequent looks (Jennison & Turnbull, 2013). To achieve the desired results, medical doctors, researchers and patients should work together with a care and royalty when clinical trials are finalized.

In general, clinical trials are classified as four consecutive phases (Friedman *et al.,* 2010). The trial design for each phase is a complex process and usually requires a close collaboration among academic institutions, medical centers or hospitals, pharmaceutical companies, public organizations and regulatory agencies. Most studies focus on human pharmaceuticals (Savci, 2016). In phase I trials, pharmaceutical operators are applied to humans for the first time (O'Quigley & Chevret, 1991). The objective of these trials is to determine any drug tolerance and interaction description properties of pharmacokinetics and to identify dosage and side effects. The main objective of a phase I trial is to understand drug tolerance in volunteers. To do so, the maximum tolerated dose (MTD) must be determined. The MTD is the highest dose of a drug which does not cause any unacceptable side effects. Several dose escalation methods are used in order to determine the MTD. Those methods fall into two classes: rule-based designs and model-based designs. The most popular method employed by scientists is the 3+3 design, one of the rule-based designs (Hansen *et al*., 2014). The main advantage of this design is that it is simple to implement.

Instead of traditional methods like 3+3 design and the Continual Reassessment Method (CRM), new phase I clinical trial methods have been developed in recent years. These include Bayesian Optimal Interval Designs (BOIN), Modified Toxicity Probability Interval Method (mTPI), Bayesian Model Averaging with Continual Reassessment Method (BMA-CRM), a Bayesian Interval Dose-Finding Design Addressing Ockham's Razor (mTPI-2), and Bayesian Continual Reassessment Method (BCRM).

There are several studies in the literature comparing phase I dose finding methods. Some studies are focused on a comparison of the rule and model based designs. Others are focused on a theoretical window. In general, a comparison of old and new dose finding methods are uncomprehensive. According to Hansen et al. (2014), 3+3 design is appropriate when the toxicity of a drug is uncertain or narrow. On the other hand, the model-based designs perform better than 3+3 design if the expected toxicity is low. It can be seen from their findings, there is no single phase I dose escalation method that produces the best results in all circumstances. Iasonos et al. (2008) conducted a study comparing CRM-based methods with the 3+3 method. They compared three

parameters: total sample size, the number of patients needed to reach the MTD, and trial duration. They concluded that CRM-based methods perform better than 3+3 design in terms of accuracy and optimal dose allocation. However, they did not include several alternative CRM-based designs in their study. Jaki et al. (2013) studied three classes of dose-escalation designs in a non-statistical window. They compared the strengths and weaknesses of rule-based designs, Bayesian model-based designs, and Bayesian curve-free methods. They concluded that rule-based designs are not recommended. In addition, they suggested that Bayesian model-based designs are appropriate if previous information is available. Paoletti et al. (2015) cameto similar conclusions, claiming that model-based methods outperform the 3+3 design in terms of selecting the correct dose level as MTD. They found that in order to adapt new objectives, model-based designs can be significantly enriched.

This study compares the most used phase I dose finding methods to determine which one performs the best. The content of this paper is as follows: The next section presents the dose escalation methods. The scenarios and simulation studies are introduced in secti on 3. The comparison of methods is also given in section 3. The last section gives the conclusion of the study.

2. Dose escalation methods

There has been an increasinginterest in phase I clinicaltrials. Thus, many phase I dose escalation studies are developed in literature. As stated, some phase I clinical trials are based on model selection while others are based on an algorithm. In this paper, both approaches are considered. The most popular choice of the phase I dose finding method among clinicians in clinical trials is 3+3 design (Ji &Wang, because it is simple and easy to implement. The 3+3 design is a rule-based design that proceeds with cohorts of three patients. DLT (dose limiting toxicity) is the most important factor in the assessment of doses. In this method, the first cohort is treated at a starting dose level (lowest dose level). The other subjects are enrolled in cohorts with increasing dose levels. The research continues with higher doses, depending on assessments of the previous doses.

A modified toxicity probability interval (mTPI) is a method that proposes dose-finding-decision rules based on the unit probability mass (UPM). This method is an improved version of the toxicity probability interval method (TPI), and it uses beta-binomial hierarchical model (Ji *et al*., 2010). The UPM is used to determine decision rules of three intervals corresponding to low, high and proper dosing in terms of toxicity (Ji & Wang, 2013). Although the mTPI allows researchers to understand the decisions before the trials start, some decision rules are debated in practice (Guo *et al.,* 2016). To prevent such debates, Yang *et al.* (2015) proposed an ad-hoc remedy that allows decision rules in the mTPI design. However, there is a lack of solid statistical justification, so it cannot be properly assessed. To handle this problem, mTPI-2 was developed. It solves the undesirable issue in the current decision under the mTPI (Guo *et al.*, 2016).

In algorithm-based dose finding methods, modeling information from the other doses is not used. In other words, it only considers the data observed from current dose level. A model-based dose finding method assumes a precise parametric model for the dose-toxicity curve. CRM, first proposed by O'Quigley *et al.* (1990), is one of the model-based approaches for dose-finding methods in drug development. CRM particularly links the likelihood of toxicity at each dose level with a pre-determined toxicity probability via a one-parameter model. While the toxicity data are collected, the CRM continually updates the estimation of the toxicity probabilities of all the doses. Each new patient cohort is transferred to the most appropriate dose based on the updated toxicity probabilities, and the MTD occurs when the entire sample size is used.

In general, toxicity is assumed to be monotonically increasing depending on the dose. Assume is a set of toxicity probabilities of a pre-determined K set of doses. These probabilities are often known as the skeleton of CRM (Wages *et al.*, 2013). Let be the predetermined possibility of target toxicity. The power-dose toxicity model for CRM in which j= 1….K is as follows.

(1)

Here, are unknown parameters. The CRM can be in different model constructions. For example, the equation of a logistic regression model, which has a *c* fixed-point, is given in the following CRM model.

where *c* is usually set to 3, and is the standard dose level at j-*th*.

Another example is the hyperbolic tangent function given as the CRM model (Eq. 3).

Another important part of the CRM is the dose finding algorithm. In practice, patients are usually treated as three-membered cohorts for each dose. Dose escalation and reduction can be limited to only one dose level at a time. The process of finding a dose for the CRM method is as follows:

Table 1. Summary of Selected Phase 1 Dose Escalation Methods

1. Start with a prior estimate of DLT for each dose level.
2. Select a mathematical model to describe the relationship between the dose and DLT.
3. Describe any uncertainty about the model by prior distribution.
4. After each patient, update the model, and estimate the probability of toxicity for each dose level.
5. Treat the next patient at the dose whose estimate is the closest to some pre-specified target.
6. Stop when a maximum sample size is reached (O'Quigley *et al*., 1996).

Another model-based design used in phase I trials is the Bayesian Continual Reassessments Method (BCRM). CRM is used because there is a need to choose the dose for the next patient by using the posterior distribution from recruited patients. BCRM is implemented by placing a prior distribution, , on the model parameter . The posterior distribution for α after n outcome is written is as

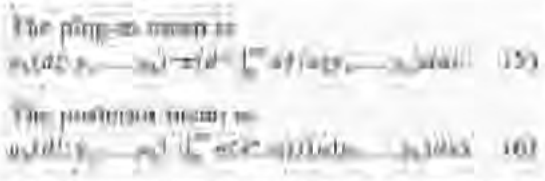
$$f(\alpha|y_1,\ldots,y_n) = \frac{f(\alpha)L(\alpha;y_1,\ldots,y_n)}{\int f(\alpha)L(\alpha;y_1,\ldots,y_n)d\alpha}, \quad (4)$$

where L()=(1-((i);α) is the likelihood. The prior distribution for α can be a number of possible, positive valued, distributions such as Gamma, Uniform or Lognormal. A

Table 2. Comparison of seven different methods for scenario 1-2 with a toxicity target 30%.

			Dose Levels					
			50mg/d	75 mg/d	100mg/d	125mg/d	150mg/d	175mg/d
Scenario 1	Method	True toxicity rate	0.02	0.06	0.12	0.20	0.30	0.40
	3+3	Probability selection	0.034	0.151	0.235	0.325	0.174	0.074
		#of patient treated	3.243	3.870	4.296	4.086	3.036	1.215
	CRM	Probability selection	0.000	0.000	0.032	0.277	0.468	0.223
		#of patient treated	3.240	3.204	4.422	7.656	7.383	4.095
	BMA-CRM	Probability selection	0.000	0.000	0.030	0.280	0.421	0.270
		#of patient treated	3.300	3.700	5.000	7.600	6.800	3.600
	BCRM	Probability selection	0.000	0.002	0.035	0.334	0.445	0.184
		#of patient treated	3.156	3.100	4.342	6.004	7.180	3.987
	mTPI	Probability selection	0.001	0.006	0.079	0.318	0.372	0.224
		#of patient treated	3.267	4.074	5.895	7.662	6.015	2.087
	mTPI-2	Probability selection	0.001	0.006	0.061	0.287	0.393	0.252
		#of patient treated	3.261	4.017	5.640	7.485	6.240	3.357
	BOIN	Probability selection	0.000	0.007	0.070	0.307	0.383	0.233
		#of patient treated	3.200	4.100	5.800	7.600	6.100	3.200
Scenario 2	Method	True toxicity rate	0.05	0.15	0.30	0.46	0.61	0.73
	3+3	Probability selection	0.217	0.268	0.428	0.057	0.004	0.000
		#of patient treated	3.966	4.147	5.978	1.776	0.375	0.018
	CRM	Probability selection	0.002	0.169	0.619	0.204	0.005	0.001
		#of patient treated	4.170	7.152	11.82	5.901	0.906	0.051
	BMA-CRM	Probability selection	0.000	0.160	0.600	0.240	0.001	0.000
		#of patient treated	4.100	7.400	12.10	5.900	0.600	0.900
	BCRM	Probability selection	0.006	0.161	0.577	0.249	0.007	0.000
		#of patient treated	3.825	6.707	11.12	5.611	0.940	0.057
	mTPI	Probability selection	0.015	0.215	0.549	0.206	0.015	0.000
		#of patient treated	4.182	8.856	11.76	4.509	0.657	0.036
	mTPI-2	Probability selection	0.015	0.219	0.548	0.200	0.018	0.000
		#of patient treated	4.284	9.159	11.33	4.569	0.819	0.036
	BOIN	Probability selection	0.001	0.235	0.559	0.182	0.001	0.000
		#of patient treated	4.200	9.300	11.10	4.700	0.700	0.000

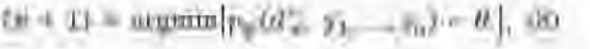
prior mean of α is represented as. Determination of the next patient's dose is generally based on point estimates of the risk of DLT. This is true for one-parameter CRM designs. However, when we need to estimate two possible points to model base dose-escalation, decisions should be made based on the following two means.



At the beginning of the trial, the point estimates used for escalation are consistent with the prior estimates of DLT written as

The dose for the (n+1)-*th* patient is the one whose point estimate is closest to the Target Toxicity Level.

So, for p=1 or 2,



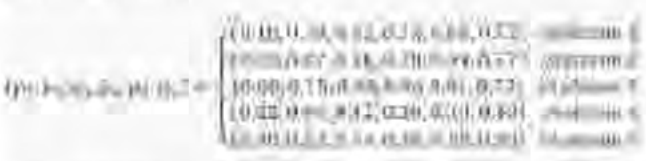
The BMA-CRM is the model-based dose finding method that aims to identify the MTD. This method pre-specifies multiple sets of mean toxicity probabilities (Yin & Yuan, 2009). As mentioned, the popular model-based dose escalation method the CRM requires predetermination of the toxicity probability at each dose. This can be discretionary and may lead to dif ferent design properties. To handle this problem, Yin & Yuan (2009) use multiple parallel CRM models each with a different set of predetermined toxicity probabilities. For each CRM model, they placed discrete probability masses. Those probability masses are used as the prior model probability. They obtained posterior probabilities by the Bayesian model averaging approach (Hoeting *et al.*, 1999). In light of this, dose increase and decrease is determined based on a target toxicity rate and dose toxicity probabilities are estimated by the BMA. Finally, the last phase I dose escalation method is BOIN. The 3+3 design is the most dominant trail based on its simplicity. The BOIN design is similarly easy to implement, and it is also flexible for choosing the target toxicity rate (Yuan *et al.*, 2016). This method minimizes the probability of inappropriate dose

assignments for patients (Liu & Yuan, 2015). So, this design has a lower risk of patient overdose. Table 1 gives a brief summary of the selected dose escalation methods.

3. Scenarios and simulation studies

A common approach for examining the properties and characteristics of model- and rule-based designs is to simulate trials according to specified scenarios. A generated result of a selected true MTD is the decision criteria for selecting the best design. In this study, we investigated the operating characteristics of seven dif ferent phase I dose escalation designs through simulation studies. Two different real life stories and six scenarios for each story were considered.

In the first story, we illustrated the proposed designs using a pharmacokinetic phase I clinical trial that investigated the safety/tolerability and the pharmacokinetic profile of SHR6390 (small molecular, oral potent, selective CDK4/6) in Chinese patients with advanced melanoma. This trial studied six different dose levels of SHR6390: 50 mg, 75 mg, 100 mg, 125 mg, 150 mg and 175 mg. The number of patients included in the study was 30. We considered the MTD as the dose with a DLT rate of 30%. We elicited five different skeletons:



In BMA-CRM, CRM and BCRM, skeletons are very important in the selection of the MTD. This is because skeletons represent dif ferent prior opinions, and they produce different MTD selections. Skeletons represent different prior guesses of the toxicity profile of a drug. The relation between the true toxicity rate and prior probabilities is very important because selection probability of MTD may increase if the prior probabilities are similar to the true toxicity probabilities. For the CRM and BCRM models, all skeletons were assigned one by one to the models, and the best performance was used for comparison. The first skeleton is for the case where toxicity starts at a high level and increases with almost the same rate. In the second skeleton, toxicity increases slowly when the dose is low,

Table 3. Comparison of seven different methods for scenario 3-4 with a toxicity target 30%.

			Dose Levels					
			50mg/d	**75 mg/d**	**100mg/d**	**125mg/d**	**150mg/d**	**175mg/d**
Scenario 3	Method	True toxicity rate	0.02	0.07	0.16	0.30	0.44	0.57
	3+3	Probability selection	0.054	0.215	0.265	0.395	0.060	0.005
		#of patient treated	3.330	4.131	4.683	5.846	1.764	0.333
	CRM	Probability selection	0.000	0.004	0.167	0.583	0.231	0.015
		#of patient treated	3.276	3.474	6.741	10.37	5.118	1.014
	BMA-CRM	Probability selection	0.000	0.000	0.170	0.562	0.250	0.030
		#of patient treated	3.300	3.900	7.100	10.20	4.700	0.800
	BCRM	Probability selection	0.000	0.007	0.177	0.580	0.224	0.012
		#of patient treated	3.154	3.200	6.245	10.01	5.111	1.007
	mTPI	Probability selection	0.001	0.016	0.237	0.530	0.194	0.022
		#of patient treated	3.297	4.488	8.328	9.507	3.762	0.618
	mTPI-2	Probability selection	0.001	0.015	0.223	0.536	0.195	0.003
		#of patient treated	3.303	4.518	8.247	9.279	3.918	0.735
	BOIN	Probability selection	0.001	0.017	0.246	0.490	0.220	0.026
		#of patient treated	3.300	4.600	8.200	9.200	4.000	0.700
Scenario 4	Method	True toxicity rate	0.05	0.10	0.15	0.20	0.30	0.40
	3+3	Probability selection	0.105	0.165	0.208	0.253	0.297	0.078
		#of patient treated	3.636	4.131	3.987	3.495	4.595	1.058
	CRM	Probability selection	0.000	0.013	0.082	0.308	0.427	0.170
		#of patient treated	3.735	3.807	5.346	7.683	6.432	2.997
	BMA-CRM	Probability selection	0.000	0.010	0.100	0.320	0.411	0.166
		#of patient treated	8.100	7.300	5.900	3.800	2.000	0.900
	BCRM	Probability selection	0.000	0.008	0.049	0.267	0.589	0.287
		#of patient treated	3.700	4.605	6.102	7.510	6.022	2.327
	mTPI	Probability selection	0.003	0.028	0.131	0.309	0.532	0.197
		#of patient treated	3.834	5.151	6.528	6.873	4.992	2.622
	mTPI-2	Probability selection	0.003	0.023	0.096	0.290	0.558	0.230
		#of patient treated	4.284	9.159	11.33	4.569	0.819	0.036
	BOIN	Probability selection	0.002	0.018	0.103	0.302	0.365	0.210
		#of patient treated	3.804	4.872	6.096	6.906	5.394	2.928

Table 4. Comparison of seven different methods for scenario 5-6 with a toxicity target 30%.

			Dose Levels					
			50mg/d	75 mg/d	100mg/d	125mg/d	150mg/d	175mg/d
	Method	True toxicity rate	0.18	0.30	0.42	0.53	0.64	0.72
Scenario 5	3+3	Probability selection	0.211	0.425	0.071	0.009	0.000	0.000
		#of patient treated	5.049	3.795	1.782	0.453	0.060	0.006
	CRM	Probability selection	0.220	0.529	0.206	0.029	0.000	0.000
		#of patient treated	10.92	11.05	6.147	1.332	0.156	0.006
	BMA-CRM	Probability selection	0.200	0.560	0.020	0.000	0.000	0.000
		#of patient treated	10.60	12.00	5.800	1.000	0.100	0.000
	BCRM	Probability selection	0.170	0.571	0.242	0.017	0.000	0.000
		#of patient treated	10.15	11.80	6.240	1.101	0.122	0.007
	mTPI	Probability selection	0.245	0.508	0.202	0.023	0.002	0.000
		#of patient treated	10.93	12.50	5.214	0.900	0.069	0.003
	mTPI-2	Probability selection	0.221	0.524	0.205	0.028	0.003	0.000
		#of patient treated	11.07	11.90	5.400	1.164	0.078	0.003
	BOIN	Probability selection	0.254	0.512	0.170	0.038	0.004	0.000
		#of patient treated	10.90	12.10	5.300	1.100	0.200	0.000
	Method	True toxicity rate	0.01	0.30	0.55	0.65	0.80	0.95
Scenario 6	3+3	Probability selection	0.105	0.165	0.208	0.263	0.257	0.078
		#of patient treated	3.636	4.131	3.987	3.495	2.595	1.058
	CRM	Probability selection	0.074	0.767	0.158	0.001	0.000	0.000
		#of patient treated	6.312	15.74	7.047	0.837	0.066	0.000
	BMA-CRM	Probability selection	0.050	0.800	0.150	0.000	0.000	0.000
		#of patient treated	5.600	17.20	6.700	0.500	0.000	0.000
	BCRM	Probability selection	0.044	0.815	0.138	0.003	0.000	0.000
		#of patient treated	6.126	16.62	6.917	0.919	0.055	0.000
	mTPI	Probability selection	0.142	0.765	0.090	0.003	0.000	0.000
		#of patient treated	6.204	18.46	4.947	0.363	0.018	0.000
	mTPI-2	Probability selection	0.124	0.786	0.087	0.003	0.000	0.000
		#of patient treated	7.998	16.63	4.995	0.363	0.018	0.000
	BOIN	Probability selection	0.164	0.759	0.073	0.004	0.000	0.000
		#of patient treated	8.000	16.60	4.900	0.400	0.000	0.000

but increases quickly at high doses. The third skeleton starts with a low dose level and increases with high dose levels. The toxicity probabilities in the fourth skeleton are gathered more at low toxicity levels. The last skeleton is concentrated in a narrow range where the toxicity probability starts at 0.2 and ends at 0.3. Tables 2 to 4 show the simulation results of the first two scenarios for the 3+3 design, CRM, BMA-CRM, BCRM, mTPI, mTPI-2, and BOIN. The probability selection of MTD and the number of treated patients are given in the tables. We carried out 10,000 simulations for each scenario.

In the first scenario, the fifth dose was the MTD. However, the 3+3 design had the lowest selection percentage of 17.4% and selected the fourth dose as the MTD with a percentage of 32.4%. On the other hand, the BMA-CRM selected the MTD with 42.1%, and a selection of BCRM performed slightly better than the BMA-CRM with 44.5%. The CRM performed better than the other methods for the selection of the true MTD with 46.8%. In the second scenario, the third dose was the MTD. The worst selection of the MTD was made by 3+3 design with 42.8%. The MTD selection percentage using the CRM was the best among the other designs. The BMA-CRM was the second best. Scenario 3 had the MTD at the fourth dose level, while the MTD selection using the CRM and BCRM had the best results with almost similar percentages. The BMA-CRM and mTPI-2 performed well, with MTD selection probabilities of 56.2% and 53.6%, respectively. In addition, the mTPI produced an almost similar MTD selection probability with the mTPI-2. In scenario 4, the BCRM and mTPI-2 produced the best MTD selection probabilities (58.9% and 55.8%). The BMA-CRM and CRM did not perform well in this scenario. In scenarios 5 and 6, the BCRM performed better than other designs with values of 57.1% and 81.5%, respectively. Other designs, except the 3+3, produced the similar selection percentage of the MTD.

From these findings, we can say that the MTD selection percentage of the first scenario, where the dose levels start low and increase gradually to a medium dose level, is not high for all designs. The selection of skeleton is very important for the CRM and BCRM. These designs can perform with the lowest selection percentage if the choice of the skeleton is inappropriate. The BMA-CRM and mTPI-2 performed very similar to the CRM and BCRM. The 3+3

design is very simple and easy to implement. However, this design had the lowest selection percentage for all scenarios.

In the second story, we applied seven designs to a phase I Reverse Genetic reassortant H9N2 influenza vaccine study conducted at Nanotherapeutics, Inc. The aim of the study was to identify the optimal dose level of a reverse genetic reassortant H9N2 influenza vaccine for further product development. This clinical trial studied six dif ferent dose levels of a Reverse Genetic reassortant H9N2 pandemic influenza vaccine in healthy subjects aged 18 to 49 years of age. Dosages were 3.75 μg, 7.5 μg, 15 μg, 30 μg, 45 μg, or 60 μg. The number of patients included in the study was 21. We considered the MTD as the dose with a DLT rate of 25%. Four different skeletons were elicited.

The same skeleton selection as in story one is used in story two. Since the skeletons represent different prior opinions, it is important to choose the right skeleton for the CRM and BCRM. The first skeleton starts at a low toxicity level and increases synchronously. The toxicity probabilities in the second skeleton changed more between 0.06 and 0.39 (low toxicity levels). In the third skeleton, toxicity increases slowly in all dose levels.

The fourth skeleton starts with a low dose level but increases quickly at high doses.The simulation results ofthe second illustration for seven different designs (3+3 design, CRM, BMA-CRM, BCRM, mTPI, mTPI-2 and BOIN) are given in Tables 5 to 7. We list the true toxicity probability in the first row. The probability selection of the MTD and the number of the treated patie nts are given in the tables. However, the comparison is made with the probability selection of the MTD. The target toxicity probability was 25%. The sample size was 21, and 10,000 simulations are carried out for each scenario.

In scenario 1, the third dose was the MTD, and all designs selected the MTD with very similar probabilities. In particular, the CRM had the highest selection percentage of 31.8% for the MTD and mTPI had the lowest selection percentage of 26.3%. The selection probabilities for MTD were very low for designs. In scenario 2, sixth dose level was the MTD, and the MTD selection percentage using the

Table 5. Comparison of seven different methods for scenario 1-2 with a toxicity target 25%.

			Dose Levels					
			3.75mg/d	7.5mg/d	15mg/d	30mg/d	45mg/d	60mg/d
Scenario 1	Method	True toxicity rate	0.15	0.20	0.25	0.30	0.35	0.40
	3+3	Probability selection	0.233	0. 265	0.283	0.103	0.034	0.017
		#of patient treated	4.611	4.921	5.730	1.662	0.816	0.240
	CRM	Probability selection	0.150	0.304	0.318	0.158	0.050	0.011
		#of patient treated	9.672	9.225	6.432	3.132	1.080	0.303
	BMA-CRM	Probability selection	0.110	0.231	0.280	0.172	0.101	0.066
		#of patient treated	8.100	7.300	5.900	3.800	2.000	0.900
	BCRM	Probability selection	0.122	0.259	0.320	0.167	0.067	0.065
		#of patient treated	9.245	8.977	6.511	3.444	1.214	0.877
	mTPI	Probability selection	0.198	0.327	0.263	0.136	0.041	0.016
		#of patient treated	1.615	1.889	1.478	0.788	0.225	0.091
	mTPI-2	Probability selection	0.173	0.336	0.273	0.128	0.056	0.015
		#of patient treated	1.661	1.775	1.449	0.806	0.295	0.121
	BOIN	Probability selection	0.071	0.201	0.280	0.233	0.134	0.073
		#of patient treated	7.002	8.208	7.211	4.504	2.001	0.806
Scenario 2	Method	True toxicity rate	0.01	0.02	0.03	0.04	0.1	0.25
	3+3	Probability selection	0.004	0.011	0.014	0.091	0.367	0.510
		#of patient treated	3.099	3.207	3.282	3.528	4.518	3.804
	CRM	Probability selection	0.000	0.000	0.003	0.012	0.229	0.756
		#of patient treated	3.123	3.285	3.414	3.675	5.874	10.62
	BMA-CRM	Probability selection	0.000	0.000	0.000	0.001	0.007	0.921
		#of patient treated	3.100	3.200	3.200	3.500	4.500	12.50
	BCRM	Probability selection	0.000	0.000	0.004	0.030	0.098	0.868
		#of patient treated	3.100	3.200	3.450	3.987	4.952	11.24
	mTPI	Probability selection	0.001	0.000	0.003	0.020	0.245	0.731
		#of patient treated	3.228	3.333	3.738	3.894	6.282	9.528
	mTPI-2	Probability selection	0.001	0.000	0.003	0.023	0.290	0.683
		#of patient treated	3.282	3.483	3.738	4.527	6.813	8.157
	BOIN	Probability selection	0.000	0.002	0.004	0.027	0.278	0.692
		#of patient treated	3.300	3.500	3.700	4.600	6.600	8.200

Table 6. Comparison of seven different methods for scenario 3-4 with a toxicity target 25%.

			Dose Levels					
			3.75mg/d	**7.5mg/d**	**15mg/d**	**30mg/d**	**45mg/d**	**60mg/d**
	Method	True toxicity rate	0.25	0.35	0.40	0.50	0.60	0.70
Scenario 3	3+3	Probability selection	0.352	0.127	0.052	0.005	0.001	0.000
		#of patient treated	5.106	2.853	1.131	0.339	0.036	0.003
	CRM	Probability selection	0.556	0.245	0.052	0.006	0.000	0.000
		#of patient treated	6.815	7.140	1.938	0.342	0.036	0.003
	BMA-CRM	Probability selection	0.450	0.170	0.050	0.001	0.000	0.000
		#of patient treated	14.20	6.000	2.300	0.500	0.100	0.000
	BCRM	Probability selection	0.552	0.302	0.116	0.025	0.004	0.001
		#of patient treated	6.755	6.441	2.114	1.482	0.004	0.003
	mTPI	Probability selection	0.432	0.241	0.063	0.007	0.000	0.000
		#of patient treated	4.473	2.582	0.718	0.128	0.011	0.000
	mTPI-2	Probability selection	0.479	0.204	0.042	0.010	0.000	0.000
		#of patient treated	4.656	2.300	0.694	0.157	0.013	0.000
	BOIN	Probability selection	0.612	0.204	0.051	0.080	0.000	0.000
		#of patient treated	7.002	8.208	7.211	4.504	2.001	0.806
	Method	True toxicity rate	0.19	0.21	0.23	0.25	0.27	0.29
Scenario 4	3+3	Probability selection	0.239	0.154	0.123	0.079	0.046	0.005
		#of patient treated	4.725	3.405	2.289	1.488	0.906	0.420
	CRM	Probability selection	0.235	0.262	0.232	0.135	0.064	0.037
		#of patient treated	11.79	8.043	5.232	2.514	1.167	0.549
	BMA-CRM	Probability selection	0.180	0.170	0.150	0.132	0.101	0.120
		#of patient treated	9.200	6.400	4.600	3.000	1.700	1.400
	BCRM	Probability selection	0.218	0.225	0.239	0.134	0.064	0.120
		#of patient treated	11.24	8.001	5.121	2.501	1.154	0.500
	mTPI	Probability selection	0.244	0.289	0.181	0.113	0.058	0.031
		#of patient treated	13.10	8.349	4.569	1.983	0.912	0.351
	mTPI-2	Probability selection	0.241	0.281	0.173	0.131	0.064	0.041
		#of patient treated	13.05	8.043	4.482	2.238	1.005	0444
	BOIN	Probability selection	0.264	0.250	0.201	0.124	0.089	0.041
		#of patient treated	12.81	7.601	4.900	2.500	1.200	0.500

BMA-CRM was the best design among others. The selection percentage of MTD for all designs was very high in this scenario comparison to the other scenarios.The reason might be the true toxicity rates because it started with low toxicity rate and increased significantly at the last toxicity rate.

Scenario 3 had the MTD at the first dose level. The CRM and BCRM were the best designs in this scenario. The 3+3 design had the lowest selection percentage of 35.2%. In this scenario, we implemented a safety rule, and in all of the designs, the trials could be terminated early. Surprisingly, the selection rate for scenario 4 was very low for all the designs. Although the fourth dose was the MTD, all designs selected a wrong dose level for this scenario. All the methods had the wrong selection of the MTD when the true toxicity rate was very slightly changed. In scenario 5, the selection percentage of the MTD for all designs was quite close, but the CRM had the highest selection rate at 45.5%. In the last scenario, the BMA-CRM was very robust. It produced the best MTD selection percentage among the other designs.

4. Conclusion

Bringing new proposed designs to phase 1 drug trials is fundamental to drug development in the initial clinical testing stage. Dose escalation methods are very important for the selection of the MTD. In this study, we neither support nor denigrate any design. Rather, we examined the properties and extensive simulation results of seven different methods.

Overall, twelve different scenarios and two different stories were considered. The model-based dose escalation designs (CRM, BCRM, and BMA-CRM) reached the MTD with similar selection percentages. In almost all simulations, the 3+3 design had the lowest selection percentage of the MTD. In the fourth scenario of the second story, where the true toxicity rate was changed very slightly, wrong dose level was selected as the MTD in all the designs. Therefore, a simulation result where the toxicity rates changed very slightly was the most dramatic one among other simulation results because selection percentage of the MTD for other simulations was correct. Thus, all designs should be checked before application in cases when the true toxicity rate is changed very slightly.

In our simulation study, the BMA-CRM, CRM and BCRM had the best performance results when compared

to other designs. The BMA-CRM design requires multiple skeletons in order to cover different scenarios. The BMA-CRM performs well if one of the skeletons corresponds to the true toxicity probabilities. For scenarios in which the toxicity probability started at a very low dose level and increase slightly until the MTD (MTD is the last dose level), BMA-CRM performed much better than the other designs. This design should be considered if the researcher has a similar scenario. When the number of patients decreased from 30 to 21 in the second story, other methods appeared to be the best design for the selection of MTD. For example, when the true toxicity rate started as the MTD and increased gradually to the high dose level, the BOIN selected the MTD with highest probability selection. Hansen *et al.* (2014) found that the 3+3 design is appropriate when the toxicity of a drug is uncertain or narrow. However, the 3+3 design appeared to be the worst design in our study. Our findings for the BMA-CRM design are similar to the study by Yin & Yuan (2009). The BMA-CRM performs well, if one of the skeletons is similar to the true toxicity rate. We found similar results to the investigation by Paoletti *et al.* (2015), claiming that model-based methods outperform the 3+3 design in terms of selecting correct dose level as the MTD.

Overall, the CRM had the best selection percentage of the MTD among the designs. However, in some cases, the BMA-CRM, BCRM, BOIN and mTPI-2 produced better results than CRM. In contrast, even though the traditional 3+3 design is very simple and easy to apply, its results were the worst in terms of performance when compared to the other designs.

In general, model-based designs produced better results. However, continual modeling by a professional is necessary. These designs can be complex for non-statisticians and might need statistical support (W ong *et al.*, 2016). In contrast, 3+3 design does not require modeling and it offers protective dose escalation for candidate drugs. However, patients may be treated at sub-therapeutic doses, and it may not be appropriate for molecularly targeted agents.

In conclusion, more reliable and applicable results for phase I dose escalation trials are produced by the BMA-CRM, CRM and BCRM designs in our study. The model-based designs performed much better than the rule-based designs.

Table 7. Comparison of seven different methods for scenario 5-6 with a toxicity target 25%.

			Dose Levels					
			3.75mg/d	7.5mg/d	15mg/d	30mg/d	45mg/d	60mg/d
	Method	True toxicity rate	0.01	0.10	0.20	0.25	0.30	0.35
Scenario 5	3+3	Probability selection	0.093	0.184	0.242	0.297	0.109	0.074
		#of patient treated	3.336	4.524	4.365	3.069	2.010	0.813
	CRM	Probability selection	0.001	0.066	0.307	0.455	0.195	0.076
		#of patient treated	3.390	5.934	9.393	10.96	3.345	1.338
	BMA-CRM	Probability selection	0.000	0.060	0.210	0.400	0.200	0.230
		#of patient treated	8.100	7.300	5.900	10.80	2.000	0.900
	BCRM	Probability selection	0.003	0.076	0.247	0.374	0.154	0.146
		#of patient treated	3.356	6.100	8.541	9.604	2.987	1.524
	mTPI	Probability selection	0.005	0.105	0.383	0.387	0.148	0.072
		#of patient treated	3.423	7.305	10.37	10.52	2.316	1.008
	mTPI-2	Probability selection	0.005	0.134	0.370	0.363	0.166	0.062
		#of patient treated	4.458	8.343	8.796	9.815	2.301	0.987
	BOIN	Probability selection	0.002	0.041	0.204	0.362	0.253	0.208
		#of patient treated	3.300	5.600	7.700	9.750	4.300	2.400
	Method	True toxicity rate	0.02	0.04	0.08	0.12	0.17	0.25
Scenario 6	3+3	Probability selection	0.012	0.071	0.134	0.193	0.232	0.351
		#of patient treated	3.183	3.525	3.915	3.975	3.669	4.472
	CRM	Probability selection	0.000	0.003	0.031	0.155	0.341	0.477
		#of patient treated	3.261	3.717	4.668	5.649	6.510	6.195
	BMA-CRM	Probability selection	0.000	0.000	0.010	0.050	0.180	0.766
		#of patient treated	3.300	3.500	4.000	4.700	5.300	9.700
	BCRM	Probability selection	0.000	0.003	0.004	0.119	0.161	0.677
		#of patient treated	3.275	3.705	4.117	5.202	5.941	6.874
	mTPI	Probability selection	0.001	0.004	0.051	0.208	0.281	0.455
		#of patient treated	3.465	4.864	5.628	6.231	5.598	5.214
	mTPI-2	Probability selection	0.001	0.002	0.058	0.185	0.337	0.417
		#of patient treated	3.588	4.386	5.733	6.129	5.391	4.773
	BOIN	Probability selection	0.001	0.070	0.055	0.168	0.364	0.406
		#of patient treated	3.600	4.500	5.700	6.100	5.400	4.800

References

Friedman, L.M., Furberg, C. & Demets, D.L. (2010). Fundamentals of clinical trials. Springer, New York. Pp. 50-130.

Guo, W., Wang, S.J., Yang, S., Lin, S. & Ji, Y. (2016). A Bayesian interval dose-finding design addressing Ockham's razor: mTPI-2. arXiv preprint arXiv:1609.08737.

Hansen, A.R., Graham, D.M., Pond, G.R. & Siu, L.L. (2014). Phase 1 trial design: Us 3+ 3 the best? Cancer Control, **21**(3): 200-209.

Hoeting, J.A., Madigan, D., Raftery A.E. & Volinsky, C.T. (1999). Bayesian model averaging: A tutorial. Statistical science, **14**(4): 382-401.

Iasonos, A., Wilton, A.S., Riedel, E.R., Seshan, V.E. & Spriggs, D.R. (2008). A comprehensive comparison of the continual reassessment method to the standard 3+ 3 dose escalation scheme in Phase I dose-finding studies. Clinical Trials, **5**(5): 465-477.

Jaki, T., Clive, S. & Weir, C.J. (2013). Principles of dose finding studies in cancer: A comparison of trial designs. Cancer Chemotherapy and Pharmacology, **71**(5): 1107-1114.

Jennison, C. & Turnbull, B.W. (2013). Interim monitoring of clinical trials: Decision theory, dynamic programming and optimal stopping. Kuwait Journal of Science, **40**(2): 43-59.

Ji, Y., Liu, P., Li, Y. & Bekele, B.N. (2010). A modified toxicity probability interval method for dose-finding trials. Clinical Trials, **7**(6):653-663.

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, **31**(14): 1785-1791.

Liu, S. & Yuan, Y. (2015). Bayesian optimal interval designs for phase I clinical trials. Journal of the Royal Statistical Society: Series C (Applied Statistics), **64**(3): 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.

O'Quigley, J., & Chevr et, S. (1991). Methods for dose finding studies in cancer clinical trials: A review and results of a Monte Carlo study. Statistics in medicine, **10**(11): 1647-1664.

O'Quigley, J. & Shen, L.Z. (1996). Continual reassessment method: A likelihood approach. Biometrics, **52**(2): 673-684.

Paoletti, X., Ezzalfani, M. & Le Tourneau, C. (2015). Statistical controversies in clinical research: Requiem for the 3+ 3 design for phase I trials. Annals of Oncology, mdv266.

Savci, S. (2016). Dubinin-radushkevich isotherm studies of equilibrium biosorption of some veterinary pharmaceuticals by using live activated sludge. Kuwait Journal of Science, **43**(3): 142-147.

Wages, N.A., Conaway , M.R., & O'Quigley , J. (2013). Performance of two-stage continual reassessment method relative to an optimal benchmark. Clinical Trials, **10**(6): 862-875.

Wong, K.M., Capasso, A., & Eckhardt, S. G. (2016). The changing landscape of phase I trials in oncology. Nature Reviews Clinical Oncology, **13**(2): 106-117.

Yang, S., Wang, S.J., and Ji, Y. (2015). An integrated dose-finding tool for phase I trials in oncology. Contemporary clinical Trials, **45**(B): 426–434.

Yin, G., & Yuan, Y. (2009). Bayesian model averaging continual reassessment method in phase I clinical trials. Journal of the American Statistical Association, **104**(487): 954-968.

Yuan, Y., Hess, K.R., Hilsenbeck, S.G., & Gilbert, M.R. (2016). Bayesian optimal interval design: A simple and well-performing design for phase I oncology trials. Clinical Cancer Research, **22**(17): 4291-4301.

Submission : 08/02/2017
Revision : 20/03/2017
Acceptance : 19/04/2017

مقارنة شاملة لطرق المرحلة الأولى للتغير في الجرعة

افيهان أولاس[2،1،*]، فيليز كارامان[2]

[1] قسم الإحصاء، كلية العلوم، جامعة كانكيري كاراتكين، كانكيري، تركيا
[2] قسم الإحصاء، كلية العلوم، جامعة يلديز للتكنولوجيا، إسطنبول، تركيا
ef_ulas@hotmail.com *

الملخص

أصبحت التجارب الاكلينيكية من النوع الأول محببة جداً في العشر سنوات الأخيرة. وتعتبر التجارب الاكلينيكية من النوع الأول من أساسيات تطوير الأدوية وذلك لأنها تقدم التصاميم المقترحة إلى المرحلة الابتدائية للاختبارات الاكلينيكية. تم حديثاً تطوير طرق عديدة لتحديد الجرعة. لكن المقارنة بين هذه التصاميم وتلك التقليدية ليست على درجة عالية من الكثافة. تهدف هذه الدراسة إلى مقارنة طرق تحديد الجرعة الأكثر استخداماً للمرحلة الأولى وتحديد الطريقة الأفضل من بينها. ولعمل ذلك، تم تحليل قصتين حقيقيتين باستخدام المحاكاة. وجدنا أن التصميم 3+3 الأكثر استخداماً من قبل العلماء هو الأسوأ مقارنةً بالتصاميم الأخرى. في المقابل كانت التصاميم BMA-CRM، CRM و BCRM أفضل من ناحية الثقة والاستخدام لتحديد الجرعة في المرحلة الأولى.