

# I期临床试验设计比较

李 萌, 张 洵

长沙理工大学, 数学与统计学院, 湖南 长沙

收稿日期: 2023年11月15日; 录用日期: 2023年12月9日; 发布日期: 2023年12月18日

## 摘 要

I期临床可以分为基于规则、基于模型和基于模型辅助三种。选取上述三种设计具有代表性的设计, 分别为3 + 3、CRM和BOIN, 进行准确性和安全性评价。首先回顾三种设计的规则, 区分三种规则之间的差异性; 其次利用蒙特卡洛进行模拟, 通过数值结果比较设计的准确性与安全性; 最后对三种设计进行讨论。

## 关键词

3 + 3, 持续再评估方法, 贝叶斯最优区间设计

# Comparison of Phase I Clinical Trials

Meng Li, Xun Zhang

School of Mathematics and Statistics, Changsha University of Science & Technology, Changsha Hunan

Received: Nov. 15<sup>th</sup>, 2023; accepted: Dec. 9<sup>th</sup>, 2023; published: Dec. 18<sup>th</sup>, 2023

## Abstract

Phase I Clinical Trials can be divided into rule-based, model-based and model-assisted design three types. Select the above three representative designs, namely 3 + 3, CRM and BOIN, to evaluate the accuracy and safety. Firstly, review the three design rules and distinguish the differences among them. Secondly, Monte Carlo simulation is used to compare the accuracy and safety of the design through numerical results. Finally, three designs are discussed.

## Keywords

3 + 3, CRM, BOIN

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

I 期临床是新药开发首次进入人体的试验, 目的是探究人体对于新药的耐受程度和药代动力学参数, 最终确定人体最大耐受剂量(MTD)。研究者们已经探出许多设计来寻找 MTD, 其中设计可分为三种类型, 基于规则、基于模型和基于模型辅助的设计[1] [2]。基于规则的设计会根据动物试验得出的临床经验以及考虑人体安全, 在试验之前就预先给出目标事件(例如, 剂量限制毒性), 然后将患者分配到各个剂量水平。基于模型的设计根据描述剂量 - 毒性关系的模型对目标毒性水平的估计来定义 MTD, 剂量 - 毒性曲线会根据累积的试验数据实时更新, 便于推荐下一组病人的“最佳剂量”。基于模型辅助的设计是近年来研究者新提出的一种设计, 具有与上述两种设计相似的操作特性。选取以上三类设计中具有代表性的设计进行操作特性比较, 因为多数新型 I 期临床试验设计都是以它们为基础进行改良的。其中基于规则设计最典型的传统 3 + 3 设计[3], 因此衍生出许多以 3 + 3 为基础的改良设计, 由 Liu Meizi 等人提出的 i3 + 3 设计[4], Simon 为改进 3 + 3 设计提出的加速滴定设计(ATD)等[5]。相比之下, 基于模型最典型的方法是 O'Quigley 等人提出的持续重新评估法(CRM) [6] [7] [8]。还有许多基于模型辅助的方法, 如 Yuan Ying 提出的 BOIN [9]设计, 该设计结合了基于规则设计的简单性和基于模型设计相似的操作特性。这些新颖的一期临床设计的发展为从业者提供了一系列工具进行更灵活和有效的第一阶段试验。

本文将在第二节中回顾 3 + 3、CRM 和 BOIN 设计的规则, 区别三种典型设计规则的差异性; 在第三节中利用 R 软件进行蒙特卡洛模拟, 通过四种指标对模拟的结果进行比较分析; 第四节对三种设计进行讨论。

## 2. 试验设计

### 2.1. 3 + 3 设计

传统 3 + 3 设计是目前应用最广泛的临床试验设计之一, 3 + 3 设计确定 MTD 的毒性概率通常需要小于 33%, 入组的 3 名病人首先需要接受研究人员事先指定的起始剂量, 假设  $j$  为最先纳入的 3 名病人所接受的剂量等级, 那么剂量升降级规则如下:

假设当前 3 名病人接受治疗的剂量等级为  $j$ :

- (i) 如果 0/3 名病人出现 DLT, 则剂量过低, 升至下一个较高剂量  $j+1$ ;
- (ii) 如果至少 2/3 名病人出现 DLT, 则降低剂量水平至  $j-1$  继续研究;
- (iii) 如果 1/3 名病人出现 DLT, 则另外增加 3 名病人, 在  $j$  剂量等级下接受相同的治疗。

在当前的 6 名病人中:

- (i) 如果 1/6 名病人出现 DLT, 则剂量过低, 升至下一个较高剂量  $j+1$ ;
- (ii) 如果 2/6 名病人出现 DLT, 将上一剂量  $j-1$  确定为 MTD;
- (iii) 如果至少 3/6 名病人出现 DLT, 则降低剂量等级为  $j-1$  继续进行研究。

### 2.2. CRM 设计

CRM 最初的模型是单参数模型, 是基于模型中具有代表性意义的设计, 事先假设药物毒性随着剂量的增大而单调递增, 其模型为:

$$\pi_j(\alpha) = p_j^{\exp(\alpha)} \quad j=1, \dots, J$$

$\pi_j(\alpha)$  代表在剂量等级下的真实毒性概率, 试验开始前确定,  $p_j$  称之为骨架,  $\alpha$  为未知参数,  $\alpha \sim N(0, 1.34)$ ,  $j$  为剂量等级。  $p_j$  借鉴 2009 年由 Lee 和 Cheung 提出的无差异区间方法, 由 R 包中的 `getprior()` 函数获取[10]。设置 MTD 分别位于六个剂量下的不同位置, 因此可以得出 6 组不同的骨架, 由于连续两个剂量之间的毒性概率间距相等, 仅选取 MTD 位于第三个剂量的骨架, 另外还任意选取了两组骨架, 选择其中一组进行试验模拟, 探究两组不同的骨架对 CRM 模型的影响。该模型能够不断更新未知参数和在每一个剂量下的药物剂量毒性概率, 与真实剂量 - 毒性曲线进行拟合, 直到耗尽最大样本量, 最终选取后验概率最接近 DLT 的剂量作为 MTD。

### 2.3. BOIN 设计

BOIN 设计是基于 Bayes 方法的区间设计, 是最经典的基于模型辅助的设计, 最早是由 Liu 和 Yuan 提出的试验设计。试验开始前设定剂量上升和下降界限, 分别为  $\lambda_e$  和  $\lambda_d$ , 剂量升降级根据当前剂量下观测的 DLT 率与界限进行比较做出决定, 参考 Liu 和 Yuan 文章中的  $\phi_1 = 0.6p_T$  和  $\phi_2 = 1.4p_T$ ,  $\lambda_e$  和  $\lambda_d$  可由下式表示:

$$\lambda_e = \frac{\log\left(\frac{1-\phi_1}{1-\phi}\right)}{\log\left(\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}\right)}$$

$$\lambda_d = \frac{\log\left(\frac{1-\phi}{1-\phi_2}\right)}{\log\left(\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}\right)}$$

其中  $\phi_1$  表示剂量不足的最高 DLT 概率, 如果小于该值, 则需要增加剂量,  $\phi_2$  表示剂量过量的最低 DLT 概率, 如果大于该值, 则需要增加剂量,  $\phi$  代表目标毒性概率, 当前病人 DLT 率为  $\hat{\pi}_j$ : 当前剂量下剂量 DLT 的总人数/当前剂量下入组的所有病人。

剂量升降规则:

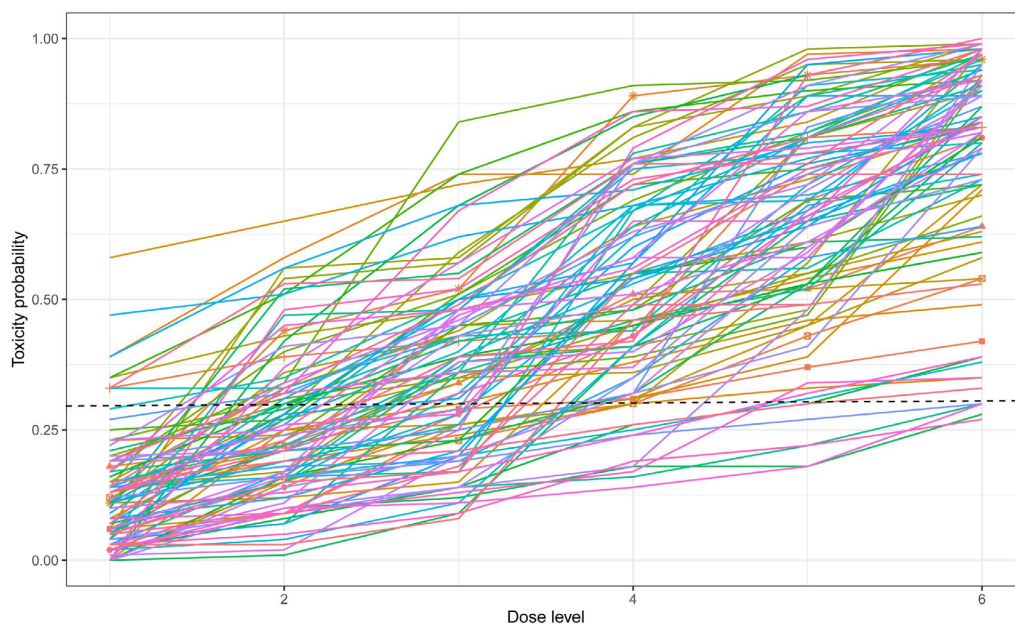
- 1) 若  $\hat{\pi}_j \leq \lambda_e$ , 则升级剂量水平至  $j+1$ ;
- 2) 若  $\hat{\pi}_j \geq \lambda_d$ , 则下降剂量水平至  $j-1$ ;
- 3) 若  $\lambda_e \leq \hat{\pi}_j \leq \lambda_d$ , 则下一入组病人继续接受当前剂量治疗。

重复上述步骤, 直到达到最大样本量, 停止试验, 决定 MTD [11]。

### 3. 蒙特卡洛模拟

利用 R 软件对三种设计进行蒙特卡洛模拟, 通过数值结果比较 3 + 3、CRM 和 BOIN 设计的操作特性。设置目标毒性概率  $\phi = 0.3$ , 共有 6 个剂量水平, 起始剂量从第一个剂量开始, 队列大小为 3, 最大样本量为 30。使用伪均匀算法生成 [0, 1] 区间的随机数, 代表真实剂量 - 毒性场景, 共生成 100 组, 最终选取 8 种模拟场景。

由图 1 可以看到多数曲线的起始剂量毒性在 0.3 以下, 但是随剂量上升, 最后剂量的毒性多数位于 0.3 以上。排除起始剂量毒性为 0 的曲线, 因为假设毒性随剂量上升而增加, 所以在起始剂量下会产生毒性反应; 排除第一剂量过毒的曲线, 如果在第一个剂量下使患者处于过毒情况, 无法保证患者安全。最终选取了 8 条适当的剂量毒性曲线。选取的 8 种场景中使毒性目标概率分散于各个剂量。具体场景如表 1:



**Figure 1.** For 100 dose-toxicity scenarios generated by pseudo-uniform algorithm, the dashed line represents the target toxicity probability of 0.3

**图 1.** 使用伪均匀算法生成的 100 个剂量 - 毒性场景, 虚线代表目标毒性概率为 0.3

**Table 1.** Eight toxicity scenarios, with bold numbers as the target toxicity probability

**表 1.** 8 种毒性场景, 加粗数字为目标毒性概率

Dose	1	2	3	4	5	6
scenario1	0.13	0.19	0.24	<b>0.30</b>	0.33	0.35
scenario2	0.12	0.16	0.23	<b>0.30</b>	0.43	0.54
scenario3	0.08	0.25	<b>0.30</b>	0.38	0.46	0.49
scenario4	0.16	<b>0.30</b>	0.39	0.45	0.53	0.59
scenario5	0.20	<b>0.30</b>	0.35	0.45	0.55	0.66
scenario6	0.02	0.08	0.14	0.26	<b>0.30</b>	0.39
scenario7	0.12	0.16	0.19	0.24	0.27	<b>0.30</b>
scenario8	0.15	0.19	0.21	0.26	<b>0.30</b>	0.33

规定目标毒性概率为 0.3, 当剂量数为 6 时, 由 R 包中的 `getprior()` 函数得到 6 组先验, 由于在实际试验中先验是由临床医生决定的, 因此再任意给出的两组先验中, 选取其中一组和 `getprior()` 函数得到的先验进行比较。探究不同先验对于 CRM 模型操作性能的影响。具体先验如表 2:

**Table 2.** Prior selection, the first six groups of prior are obtained by the `getprior()` function in the R package, and the last two groups of prior are given arbitrarily, and the bold prior is used in the simulation experiment

**表 2.** 先验选取, 前 6 组先验由 R 包中的 `getprior()` 函数获得, 后两组先验任意给出, 加粗先验用于模拟试验

	CRM 先验					
先验 1	0.300	0.423	0.537	0.633	0.708	0.765
先验 2	0.187	0.300	0.423	0.537	0.633	0.708

## Continued

先验 3	<b>0.101</b>	<b>0.187</b>	<b>0.300</b>	<b>0.423</b>	<b>0.537</b>	<b>0.633</b>
先验 4	0.046	0.101	0.187	0.300	0.423	0.537
先验 5	0.018	0.046	0.101	0.187	0.300	0.423
先验 6	0.006	0.018	0.046	0.101	0.187	0.300
先验 7	0.10	0.20	0.30	0.40	0.50	0.60
先验 8	<b>0.06</b>	<b>0.13</b>	<b>0.19</b>	<b>0.24</b>	<b>0.27</b>	<b>0.30</b>

## 指标选取及模拟结果

基于以下四种评价指标比较三种设计的准确性与安全性, 四种指标分别为:

1) MTD 正确择取率, 择取 MTD 的区间为 $[0.25, 0.33]$ 。如果 MTD 位于该区间内, 则可以选择该剂量为最大耐受剂量。

2) 在 MTD 下分配的病人概率。如果在 MTD 概率数值大, 则说明能有更多的病人在该剂量下被治疗。

3) 在 MTD 之前停止的概率。概率越大说明在 MTD 之前治疗的人数更多, 能够防止更多的病人暴露在过量剂量下。

4) 过毒人数概率。能够反映在过量剂量下暴露的病人数, 概率越大说明有更多的病人暴露在过量剂量下。

具体结果见表 3:

**Table 3.** Comparison of simulation results of three designs

**表 3.** 三种设计的模拟结果比较

	MTD 正确择取率							
	场景 1	场景 2	场景 3	场景 4	场景 5	场景 6	场景 7	场景 8
3 + 3	0.172	0.173	0.475	0.265	0.232	0.378	0.174	0.254
CRM-1	0.404	0.374	0.701	0.487	0.424	0.746	0.351	0.546
CRM-2	0.341	0.311	0.631	0.511	0.412	0.434	0.626	0.669
BOIN	0.384	0.331	0.642	0.444	0.344	0.635	0.431	0.549
	在 MTD 下分配的病人概率							
	场景 1	场景 2	场景 3	场景 4	场景 5	场景 6	场景 7	场景 8
3 + 3	0.094	0.110	0.179	0.338	0.305	0.093	0.039	0.050
CRM-1	0.195	0.231	0.311	0.356	0.312	0.199	0.062	0.090
CRM-2	0.139	0.179	0.267	0.346	0.297	0.184	0.217	0.110
BOIN	0.163	0.193	0.244	0.389	0.331	0.159	0.058	0.091
	在 MTD 之前停止的概率							
	场景 1	场景 2	场景 3	场景 4	场景 5	场景 6	场景 7	场景 8
3 + 3	0.604	0.543	0.423	0.202	0.265	0.549	0.826	0.746
CRM-1	0.552	0.500	0.321	0.173	0.240	0.506	0.862	0.757
CRM-2	0.403	0.404	0.316	0.164	0.231	0.244	0.530	0.515
BOIN	0.520	0.484	0.460	0.304	0.370	0.511	0.748	0.678

Continued

	过毒人数概率							
3 + 3	0.059	0.059	0.095	0.167	0.163	0.050	0	0.027
CRM-1	0.094	0.100	0.221	0.335	0.302	0.091	0	0.038
CRM-2	0.243	0.222	0.278	0.350	0.327	0.278	0	0.155
BOIN	0.103	0.103	0.161	0.244	0.240	0.065	0	0.038

由表 3 可以看出, 就 MTD 正确择取率而言, 在 8 个场景下, 3 + 3 设计均处于劣势, 而 CRM 和 BOIN 表现较好, 由于 CRM 先验选取的影响, 表现出不同的操作特性, 在场景一下, CRM-1 正确择取率为 0.404, CRM-2 为 0.341, 而 BOIN 为 0.384, BOIN 在场景一下优于 CRM-2, 场景二、场景三、场景五、场景六、场景七和场景八中 BOIN 优于 CRM, 而在场景四下, BOIN 为 0.444, CRM 均优于 BOIN。由数值结果可以观察到 3 + 3 能够更准确地选择 MTD, 而 CRM 由于先验的影响会造成选择的偏倚, 与 BOIN 设计相比, CRM 在某些场景下正确择取率并不高。

在 MTD 下分配的病人概率可以评估在 MTD 下被治疗的人数, 概率越大, 表明在 MTD 下被治疗的病人更多。由上表可以看出, 在场景五下, 3 + 3 分配的人数比例为 0.305, CRM-1 的概率为 0.312, CRM-2 为 0.297, BOIN 设计为 0.331, 可以看出 3 + 3 优于 CRM-2, 但数值仅相差 0.008, 相比之下, BOIN 设计则能够在 MTD 下治疗更多的病人。在其余七个场景下, 3 + 3 表现最差, 其次是 CRM, 而 BOIN 表现最优。由于 CRM 设计先验的影响, 在某些场景下表现性能劣于 BOIN。

在 MTD 之前就停止概率, 该指标可以说明能够更快地择取 MTD, 而花费更少的次数在剂量之间跳跃。由数值结果可以看出, 在场景一、场景二和场景六下 3 + 3 表现最优, 但是数值结果差值与 CRM 相差不到 0.1, 其次是 CRM, 在场景七下, CRM-1 高达 0.862, CRM-2 为 0.530, 由于先验的影响, 数值相差高达 0.332, 说明先验对于 CRM 的影响比较大, 其次是 BOIN, BOIN 在多数场景下表现与 CRM 相当。

过毒人数概率为选择大于 0.3 分配的病人百分数评价设计安全性。3 + 3 在所有场景下表现最优, 能够防止更多的病人暴露在目标毒性以上, 其次是 BOIN, 表现最差的是 CRM。虽然基于模型和基于辅助模型的设计择取 MTD 更加精确, 并且灵活性强, 但是使患者暴露在过量毒性的风险更大。

#### 4. 讨论

I 期临床试验一直以来是开展临床的关键阶段, 因此探究 I 期试验设计的准确性和安全性是有必要的。本文介绍了 3 + 3、CRM 和 BOIN 设计的规则, 并且通过模拟操作说明了设计的优劣性, 其中还对 CRM 受先验影响进行了数值结果比较, 最终结果显示, CRM 会由于先验的选取使得结果受到偏倚, 3 + 3 设计由于其规则的简单性正确择取率偏低, 而 CRM 和 BOIN 择取 MTD 的精度更高。在大部分场景下, CRM 和 BOIN 能够使得更多的病人在 MTD 下进行治疗并且能够更快地找到 MTD, 而 3 + 3 设计则会使更多的病人暴露在低剂量下。就安全性而言, 3 + 3 表现最优, 能够使更多的病人在低于毒性剂量下接受治疗。因此可由数值结果得出, 基于规则的试验设计由于其简单的剂量升级规则, 能够降低病人暴露在过量毒性下的风险, 是临床试验应用的可靠选择。基于模型的设计需要利用数学模型, 可以事先进行模拟, 对于 MTD 选择更加准确, 但是由于其程序的复杂性, 因此对实际临床试验的应用较少。基于辅助模型的设计结合了上面两种设计的优点, 近年来得到广泛的应用。三种设计各有优势, 因此需结合具体情况对试验设计进行选择。

#### 参考文献

- [1] Le Tourneau, C., Lee, J.J. and Siu, L.L. (2009) Dose Escalation Methods in Phase I Cancer Clinical Trials. *Journal of*

- 
- the National Cancer Institute*, **101**, 708-720. <https://doi.org/10.1093/jnci/djp079>
- [2] 李若冰, 何娟, 言方荣, 等. 抗肿瘤药物早期临床研究的新颖试验设计方法[J]. 中国新药杂志, 2021, 30(24): .
- [3] Storer, B. and DeMets, D. (1987) Current Phase I/II Designs: Are They Adequate. *Journal of Clinical Research and Drug Development*, **1**, 121-130.
- [4] Liu, M., Wang, S.J. and Ji, Y. (2019) The i3+3 Design for Phase I Clinical Trials. *Journal of Biopharmaceutical Statistics*, **30**, 1-11. <https://doi.org/10.1080/10543406.2019.1636811>
- [5] Simon, R., Rubinstein, L., Arbuck, S.G., Christian, M.C., Freidlin, B. and Collins, J. (1997) Accelerated Titration Designs for Phase I Clinical Trials in Oncology. *Journal of the National Cancer Institute*, **89**, 1138-1147. <https://doi.org/10.1093/jnci/89.15.1138>
- [6] O'Quigley, J., Pepe, M. and Fisher, L. (1990) Continual Reassessment Method: A Practical Design for Phase I Clinical Trials in Cancer. *Biometrics*, **46**, 33-48. <https://doi.org/10.2307/2531628>
- [7] Chevret, S. (1993) The Continual Reassessment Method in Cancer Phase I Clinical Trials: A Simulation Study. *Statistics in medicine*, **12**, 1093-1108. <https://doi.org/10.1002/sim.4780121201>
- [8] Saxena, A., Rubens, M., Ramamoorthy, V., *et al.* (2022) A Brief Overview of Adaptive Designs for Phase I Cancer Trials. *Cancers*, **14**, Article No. 1566. <https://doi.org/10.3390/cancers14061566>
- [9] Liu, S.S. and Yuan, Y. (2015) Bayesian Optimal Interval Designs for Phase I Clinical Trials. *Journal of the Royal Statistical Society*, **64**, 507-523.
- [10] Lee, S.M. and Cheung, Y.K. (2009) Model Calibration in the Continual Reassessment Method. *Clinical Trials Journal of the Society for Clinical Trials*, **6**, 227-238.
- [11] Zhou, H., Yuan, Y. and Nie, L. (2018) Accuracy, Safety, and Reliability of Novel Phase I Trial Designs Accuracy, Safety, and Reliability of Novel Phase I Designs. *Clinical Cancer Research*, **24**, 4357-4364. <https://doi.org/10.1158/1078-0432.CCR-18-0168>