

【生物与医学统计】

自适应临床试验设计的近期进展及展望

高敬雅 马 维 张绍洪 胡飞芳

【摘要】由于具有在伦理道德、有效性和灵活性等方面的优势,自适应设计在临床试验中受到越来越广泛的关注。本文首先介绍了两类主要的自适应设计,随机化罐子模型和基于序贯估计的方法。随后,本文着重介绍了自适应设计的近期进展:(1)自适应设计中的序贯监测问题;(2)自适应设计在响应变量为生存数据、复发事件数据和有序数据的试验中,以及在非劣效试验中的应用。最后我们探讨了自适应设计中尚需进一步解决的问题以及今后可能的研究方向。

【关键词】自适应设计;随机化罐子模型;双重适应性有偏硬币设计;最优分配

【作者简介】高敬雅,中国人民大学统计学院;马维,中国人民大学统计与大数据研究院(北京 100872);张绍洪,南方科技大学统计与数据科学系(广东 深圳 518055);胡飞芳,乔治华盛顿大学统计系(华盛顿 20052)。

【原文出处】《数理统计与管理》(京),2020.4.595~610

【基金项目】本文得到了国家自然科学基金(11731011,11801559)的支持。

0 引言

在传统的临床试验中,研究者通常采用双盲的平行对照设计,将受试者按照固定的概率随机分配到各试验组中,直至所有受试者分配完成。这种试验方案简单明了,便于实施,但是这种试验方案通常以相同的概率将受试者分配到治疗效果好和治疗效果差的治疗组(或安慰剂组)中,在治疗癌症、艾滋病等恶性疾病的试验中会面临伦理道德方面的挑战。由于在临床试验中,受试者通常在一段时间内序贯进入试验,研究者可以利用已经收集到的分组情况和治疗结果等信息,有计划地调整试验方案,来实现特定的目标,如减少分配到较差试验组的人数,缩短试验周期等。这种根据已有信息,合理调整试验方案的设计,称为适应性设计(adaptive design)。对适应性设计模型及应用的介绍可参见Chow和Chang^[1]及Bothwell等^[2],我国学者也曾对适应性设计的研究进展、应用及伦理性进行了介绍与探讨^[3-5]。

调整随机化分配过程的适应性设计,又称适应性随机化设计(adaptive randomization designs)。随机化是临床试验设计的一个重要原则,它是进行统计推断的基础,并且能够保证治疗组之间的可比性及减少试验过程的选择偏倚。适应性随机化设计是一类重要的临床试验设计,能够在分配过程中根据已收集到的信息调整随机分配概率,来实现伦理道德或统计检验方面的特定目标。Rosenberger和Lachin^[6]在专著中详细介绍了适应性随机化设计中的基本问题及各类适应性随机化模型。根据随机化分配过程中所利用的信息及目的的不同,Hu和Rosenberger^[7]将适应性随机化设计分为四类:(1)限制性随机化设计(restricted randomization):根据已有的分组信息调整分配概率,使各组受试者数量趋于一致;(2)协变量适应性随机化设计(covariate-adaptive randomization):根据已有的受试者分组情况及协变量数据调整分配概率,使各组的协变量分布一致;(3)自适应设计(response-adaptive ran-

domization):根据已有的分组信息及治疗结果数据调整分配概率,以实现特定的目标分配比例;(4)协变量调整—自适应设计(covariate-adjusted response-adaptive randomization, CARA):根据已有的受试者分组信息、协变量及治疗结果数据调整分配概率,在考虑受试者协变量与治疗结果相关性的基础上,分配更多的受试者到较好的治疗组中。

本文将重点介绍第三类适应性随机化设计——自适应设计。自适应设计从提出至今已有几十年的发展历史,目前研究者们已经为自适应设计建立了较为完整理论框架,并推导出了相应的渐近性质。详细内容可参见 Hu 和 Rosenberger^[7]2006 年的专著。其中,由于双重适应性有偏硬币设计(doubly-adaptive biased coin design, DBCD)具有适用于各种响应数据类型、可实现任意给定的目标分配比例及实际分配比例的方差接近方差下界等优势,最近十几年来受到了研究者的广泛的关注。本文主要介绍以 DBCD 为主的自适应试验设计的近期进展及展望:第 1 节将介绍两类自适应设计及其近年来的研究进展;第 2 节将详细介绍 DBCD 过程在不同数据类型及不同试验类型中的应用;最后在第 3 节对自适应设计进行总结,并讨论自适应设计中还未解决的问题及未来发展的方向。

1 自适应设计

根据随机化分配工具的不同,自适应设计主要可以分为随机化罐子模型(randomized urn models)和基于序贯估计的方法。本节首先阐述自适应设计的发展历程,然后分别介绍两类自适应设计的典型模型,以及如何对自适应设计进行成组序贯监测。

1.1 自适应设计的发展历程

自适应设计,又称反应变量—适应性设计^[8]或响应—自适应设计^[9],它最初的目的是根据已知的治疗结果数据,将更多的受试者分配到治疗效果较好的治疗组中。这一思想最早可追溯到 Thompson^[10]和 Robbins^[11],随后 Anscombe^[12]和 Colton^[13]根据自适应思想进行了很多拓展性的研究工作。1969 年 Zelen^[14]根据 Robbins^[11]的工作提出了著名

的非随机化的自适应设计——胜者优先原则(play-the-winner rule, PW)——根据当前受试者的治疗结果决定下一位受试者所接受的治疗。Wei 和 Durham^[15]于 1978 年将 PW 引入罐子模型,提出了随机化胜者优先模型(randomized play-the-winner rule, RPW),通过从罐子中随机抽球来决定受试者所分配的治疗组,其中罐子里各类球的组成与前面受试者的治疗结果有关,治疗效果好的治疗组对应的球数更多。Ivanova^[16]在 RPW 的基础上,提出一种新的规则来调整罐子里各类的球数目,即 drop-the-loser rule (DL)。相较于 RPW, DL 波动性更小。在自适应设计中,通常采用实际分配比例的方差来衡量设计的波动性^[17]。这一类通过从罐子中抽取球来进行随机化分配的试验过程一般称为随机化罐子模型。常见的随机化罐子模型还包括广义 Friedman 罐子模型(Generalized Friedman Urn, GFU)^[18-19]、birth and death 罐子模型(birth and death urn, BDU)^[20]以及广义 drop-the-loser 模型(generalized drop-the-loser rule, GDL)^[21]等。

随机化罐子模型的极限分配比例一般不是统计检验意义上的最优分配(optimal allocation)。在只考虑检验功效的情况下,若样本量固定, Neyman 分配的检验功效最大。为了使实际分配比例达到最优,以实现相应的目标, Melfi 和 Page^[22]独立于罐子模型的框架,提出了一个新的自适应设计,以目标分配的序贯估计为分配概率进行随机分配,又称序贯极大似然设计(sequential maximum likelihood procedure, SMLP)。这种根据参数的序贯估计进行分配的试验设计即为基于序贯估计的自适应设计。以不同的最优分配为目标分配,即可实现相应的目的。因此在实施这类自适应设计前,需要提前给定最优分配比例, Rosenberger 等^[23]和 Tymofyeyev 等^[24]对最优分配问题进行了详细的讨论。此外, Eisele^[25]及 Eisele 和 Woodroffe^[26]提出了另一类基于序贯估计的自适应设计,双重适应性有偏硬币设计(DBCD),并证明在一组严格的条件下,实际分配比例能收敛到目标分配。Hu 和 Zhang^[27]提出一族

新的 DBCD 分配函数,并在更一般的限制条件下,推导出 DBCD 过程的渐近性质。Hu 等^[28]证明,DBC-CD 过程实际分配比例的方差虽然无法达到方差下界,但两者之间的差距非常小。2009 年, Hu 等^[29]提出了有效自适应设计 (efficient randomized - adaptive design, ERADE), 该设计实际分配比例无偏且方差能够到达方差下界,即 ERADE 为一个渐近最优自适应设计。

此外,为了解决罐子模型无法实现最优分配的问题,Zhang 等^[30]于 2006 年提出了序贯估计调整的罐子模型 (sequential estimation - adjusted urn, SEU)。2011 年, Zhang 等^[31]提出的 Immigrated 罐子模型 (immigrated urn model, IMU) 几乎包含了所有常见的罐子模型,为临床试验中的罐子模型建立了一个统一的理论框架。

1.2 Immigrated 罐子模型 (IMU)

在 K 个治疗组的临床试验中,假设罐子中有 $K+1$ 类球,标记为 0 的球为附加球,标记为 $1, 2, \dots, K$ 的球分别对应 K 个治疗组,罐子的初始组成为 $\mathbf{Y}_0 = (Y_{0,0}, Y_{0,1}, \dots, Y_{0,K})'$,即罐子中有 $Y_{0,k}$ ($k=0, 1, \dots, K$) 个 k 类球。当第 i 位受试者等待分配时,罐子中各类球的组成为 $\mathbf{Y}_{i-1} = (Y_{i-1,0}, Y_{i-1,1}, \dots, Y_{i-1,K})'$,从罐子中无放回地随机抽取一个球。若抽到附加球,则将附加球放回,并向罐子中分别添加 $a_{i-1,k} \geq 0$ 个 k 类球 ($k=1, 2, \dots, K$),重复这一过程,直到抽到非附加球。若抽到第 j ($\neq 0$) 类球,则对第 i 位受试者进行第 j 组治疗并观察其治疗结果 $\xi_{i,j}$,然后向罐子中分别添加 $D_{j,k}(i)$ 个 k 类球 ($k=1, 2, \dots, K$),其中 $D_{j,k}(i)$ 为响应变量 ξ_i 的函数。假设响应变量 $\boldsymbol{\xi}_i = (\xi_{i,1}, \dots, \xi_{i,K})'$ 的分布参数为 $\boldsymbol{\theta}$,且参数 $\boldsymbol{\theta}$ 基于前 $i-1$ 个治疗结果的估计为 $\hat{\boldsymbol{\theta}}_{i-1}$ 。在 IMU 中, $\mathbf{D}_i = (D_{j,k}(i))_{j,k=1}^K$ 的元素可以为负数,表示从罐子中取出球。假设在 IMU 模型中, $a_{i,k}$ 可以是参数估计的函数 $a_{i,k} = a_k(\hat{\boldsymbol{\theta}}_i)$,并且罐子中的球数及每步添加的球数都可以是非整数的实数。

当 $\mathbf{Y}_0, a_{i,k}$ 和 $D_{j,k}(i)$ 的取值情况不同时,IMU 分别对应不同的罐子模型,例如:

1. RPW: $K=2$, 令 $Y_{0,0} = 0$ 或 $a_{i,k} = 0, \forall i, k$, 当第 i 位受试者被分配到第 j 个治疗组时,若治疗结果为成功,则 $D_{j,j}(i) = 2, D_{j,k}(i) = 0, k \neq j$, 否则 $D_{j,k}(i) = 1, k = 1, \dots, K$;

2. GFU: 令 $Y_{0,0} = 0$ 或 $a_{i,k} = 0, \forall i, k, D_{j,k}(i)$ 为响应变量 ξ_i 的函数;

3. BDU: 令 $a_{i,k} = 1, k = i, \dots, K$, 当第 i 位受试者被分配到第 j 个治疗组时,若治疗结果为成功,则 $D_{j,j}(i) = 2, D_{j,k}(i) = 0, k \neq j$, 否则 $D_{j,k}(i) = 0, k = 1, \dots, K$;

4. DL: 令 $a_{i,k} = 1, k = i, \dots, K$, 当第 i 位受试者被分配到第 j 个治疗组时,若治疗结果为成功,则 $D_{j,j}(i) = 1, D_{j,k}(i) = 0, k \neq j$, 否则 $D_{j,k}(i) = 0, k = 1, \dots, K$;

5. GDL: 令 $\mathbf{a}_i = C\boldsymbol{\rho}(\hat{\boldsymbol{\theta}}_i)$, 其中 $\mathbf{a}_i = (a_{i,1}, \dots, a_{i,K})$, C 为给定常数且 $\boldsymbol{\rho}(\hat{\boldsymbol{\theta}}_i)$ 为目标分配比例 $\boldsymbol{p}(\boldsymbol{\theta})$ 的估计, $D_{j,k}(i)$ 为响应变量 ξ_i 的函数;

6. SEU: 令 $Y_{0,0} = 0$ 或 $a_{i,k} = 0, \forall i, k, D_{j,k}(i)$ 为参数估计 $\hat{\boldsymbol{\theta}}_{i-1}$ 的函数;

IMU 的实际分配比例可以收敛到任意给定的比例,并服从渐近正态分布,其详细的渐近性质可参见 Zhang 等^[31]。

1.3 双重适应性有偏硬币设计 (DBCD)

Eisele^[25], Eisele 和 Woodroffe^[26] 提出的 DBCD 能够使各个治疗组的人数比例快速地收敛到目标比例。DBCD 在分配受试者时,分配概率不仅依赖于目标分配比例的估计值,还与目前的实际分配比例有关。在 K 个治疗组的临床试验中,记 $\boldsymbol{\theta}$ 为响应变量的分布参数,假设当第 $i-1$ 个受试者分配完成后,各治疗组人数为 $N(i-1) = (N_1(i-1), \dots, N_K(i-1))$,目标分配比例 $\boldsymbol{\rho}(\boldsymbol{\theta})$ 的估计为 $\hat{\boldsymbol{\rho}}(i-1) = (\rho_1(\hat{\boldsymbol{\theta}}_{i-1}), \dots, \rho_K(\hat{\boldsymbol{\theta}}_{i-1}))$,则 DBCD 分配过程为:

1. 试验开始时,将 Kn_0 个位受试者随机地平均分配到 K 个治疗组中接受治疗;

2. 当第 i 位受试者 ($i > Kn_0$) 进入试验时,将第 i 个受试者分配到第 j 个治疗组的概率为

$$p_{i,j} = g\left(\frac{N_j(i-1)}{i-1}, \rho_j(\hat{\boldsymbol{\theta}}_{i-1})\right), \quad j=1, \dots, K,$$

其中 n_0 为提前给定的常数。这一分配过程的统计性质主要由分配函数 $g(\cdot, \cdot)$ 决定。

Eisele 和 Woodroffe^[26] 考虑了 $K=2$ 的情况, 并给出一个分配函数的例子

$$g(x, y) = \left(1 - \left(\frac{1}{y} - 1\right)x\right)_+, \quad (1)$$

同时他们对 $g(\cdot, \cdot)$ 列出了一组较为严格的限制条件, 在这些条件下实际分配比例可以收敛到目标比例。但 Melfi 等^[32] 指出, 分配函数式(1)并不完全满足 Eisele 和 Woodroffe 提出的限制条件。

Hu 和 Zhang^[27] 提出了一族新的 DBCD 分配函数, 并将其推广到 $K \geq 2$ 个治疗组的情况中。当 $K=2$ 时, 其分配函数为

$$g(x, y) = \frac{g\left(\frac{y}{x}\right)^\gamma}{y\left(\frac{y}{x}\right)^\gamma + (1-y)\left(\frac{1-y}{1-x}\right)^\gamma}, \quad 0 < x < 1, \quad 0 < y < 1, \quad (2)$$

并且当 $x=0$ 时, $g(0, y) = 1$, 当 $x=1$ 时, $g(1, y) = 0$, 其中, $\gamma \geq 0$ 为随机化调节参数。当 $\gamma=0$ 时, 即为 SMLP 设计。 γ 越大, x 趋近于 y 的速度越快, 分配过程的随机化程度越低。此外, Hu 和 Zhang^[26] 证明了在更一般的限制条件下, 分配函数式(2)可保证实际分配比例收敛到目标比例。Eisele 和 Woodroffe^[26] 提出的分配函数式(1)满足这组一般化的限制条件。

假设当 $K=2$ 时, DBCD 过程采用分配函数式(2)。若 $\boldsymbol{\rho}(\boldsymbol{\theta}) = (\rho_1(\boldsymbol{\theta}), \rho_2(\boldsymbol{\theta}))'$ 连续且在 $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)'$ 的邻域内连续二阶可导, 响应变量 $\boldsymbol{\xi}_i = (\xi_{i,1}, \xi_{i,2})'$, $i=1, \dots, n$, 独立同分布且对任意 $\varepsilon > 0$, 有 $E \|\boldsymbol{\xi}_{1,j}\|^{2+\varepsilon} < \infty, j=1, 2$, 则

$$n^{\frac{1}{2}} \left(\frac{N_j(n)}{n} - \rho_j(\boldsymbol{\theta}) \right) \xrightarrow{D} N \left(0, \frac{1}{1+2\gamma} c_3 + \frac{2(1+\gamma)}{1+2\gamma} c_4 \right),$$

其中, $c_3 = \rho_1(\boldsymbol{\theta})\rho_2(\boldsymbol{\theta}), c_4 = (\nabla \boldsymbol{\rho}(\boldsymbol{\theta})|_{\boldsymbol{\theta}})' \mathbf{V} \nabla \boldsymbol{\rho}(\boldsymbol{\theta})|_{\boldsymbol{\theta}}, \mathbf{V} = \text{diag} \left(\frac{\text{Var}(\xi_{1,1})}{\rho_1(\boldsymbol{\theta})}, \frac{\text{Var}(\xi_{1,2})}{\rho_2(\boldsymbol{\theta})} \right)$ 。当 $K > 2$ 时, 实际分配比例和目标比例估计的渐近正态性, 可参见 Hu 和 Zhang^[27]。肖意可等^[33] 通过对实际案例的模拟研究, 说明了 DBCD 过程的可行性及优势。值得

注意的是, DBCD 的波动性与分配函数的参数 γ 有关, γ 越大 DBCD 的波动性越小^[27]。当 γ 趋于无穷时, DBCD 实际分配比例的方差能够达到方差下界, 但此时 DBCD 的分配过程不再是随机的。通常建议 γ 的取值在 $[2, 4]$ 范围内, 此时 DBCD 可以兼顾随机性和波动性之间的平衡, 并且实际分配比例的方差接近方差下界。后文中所提到的 DBCD 过程, 若非特意注明, 均表示基于 Hu 和 Zhang^[27] 分配函数式(2)的 DBCD 过程。

1.4 自适应设计的成组序贯监测

在实际临床试验中, 成组序贯设计 (group sequential design) 是目前最常用的适应性设计。监测委员会定期检查期中数据, 通过验证试验安全性, 或检验现有数据是否有足够证据证明治疗有效或无效, 来决定提前终止试验或继续进行试验。Jennison 和 Turnbull^[34] 在专著中介绍了临床试验中的成组序贯方法, 并总结了期中分析的三个方面的优势: (1) 伦理方面, 可以尽早避免受试者接受有危险的治疗; (2) 管理方面, 能够监控试验过程是否符合研究计划和前期假设; (3) 经济方面, 可减少样本量及试验花费。

Zhu 和 Hu^[35] 首次从理论方面研究了如何对自适应设计进行序贯监测。假设试验共有两个治疗组, 样本量为 n , 令 $\mathbf{T}_i = (T_{i1}, T_{i2})' = (1, 0)', i=1, \dots, n$, 表示第 i 个受试者分配到第一个治疗组, 否则 $\mathbf{T}_i = (0, 1)'$ 。令 $N_j(n) = \sum_{i=1}^n T_{ij}, j=1, 2$, 则 $N(n) = (N_1(n), N_2(n))'$ 表示两个治疗组的最终分配人数。响应变量 $\boldsymbol{\xi}_i = (\xi'_{i1}, \xi'_{i2})', i=1, \dots, n$ 独立同分布, 分布参数为 $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2)'$, 简单起见, 令 $\boldsymbol{\xi}_{ij}$ 与 $\boldsymbol{\theta}_j$ 的维度相同, 且 $\boldsymbol{\theta}_j = E(\boldsymbol{\xi}_{ij}), j=1, 2$, 否则可通过对 $\boldsymbol{\xi}_{ij}$ 进行变换, 使其满足该等式^[27,36]。若 N 为已经分配的受试者数量, 定义 $t = \frac{N}{n}, t \in [0, 1]$, 为信息量时间参数, 以便在 Skorohod 空间中研究随机化分配过程。令 $[nt]$ 表示不大于 nt 的最大整数, 则在 $N = [nt]$ 个受试者分配完成后, 可得到参数 $\boldsymbol{\theta}$ 此时的估计 $\hat{\boldsymbol{\theta}}([nt]) = (\hat{\boldsymbol{\theta}}'_1([nt]), \hat{\boldsymbol{\theta}}'_2([nt]))'$,

$$\hat{\theta}_1([nt]) = \frac{\sum_{i=1}^{[nt]} T_{i1} \xi_{i1} + \theta_{0,1}}{N_1([nt]) + 1}, \hat{\theta}_2([nt]) = \frac{\sum_{i=1}^{[nt]} T_{i2} \xi_{i2} + \theta_{0,2}}{N_2([nt]) + 1}, \quad (3)$$

其中,分母中加“1”来保证估计的连续性,分子中加“ $\theta_{0,j}$ ”可在没有受试者分配到该治疗组时用来估计参数,如令 $\theta_{0,j}$ 为某个给定的常数向量。考虑试验采用 DBCD 过程,目标分配比例为 $\rho = \rho(\theta)$,并使用 Hu 和 Zhang^[27]提出的分配函数式(2)来随机分配受试者。

在比较两个治疗组时,建立一般假设检验:

$$H_0: h(\theta_1) = h(\theta_2) \text{ vs } H_1: h(\theta_1) \neq h(\theta_2), \quad (4)$$

其中 h 是关于参数的连续函数,且在 $\theta_j, j=1,2$ 的邻域内二阶可导。若考虑在时间点 $t \in (0, 1]$ 对上述假设进行检验,可建立检验统计量

$$Z_t \left(\frac{N([nt])}{[nt]}, \hat{\theta}([nt]) \right) = \frac{h(\hat{\theta}_1([nt])) - h(\hat{\theta}_2([nt]))}{\sqrt{\widehat{\text{Var}}(h(\hat{\theta}_1([nt]))) + \widehat{\text{Var}}(h(\hat{\theta}_2([nt])))}}, \quad (5)$$

其中 $\widehat{\text{Var}}(h(\hat{\theta}_1([nt])))$ 和 $\widehat{\text{Var}}(h(\hat{\theta}_2([nt])))$ 分别为 $h(\hat{\theta}_1([nt]))$ 和 $h(\hat{\theta}_2([nt]))$ 的方差的一致估计。Zhu 和 Hu^[35]证明在一组广泛适用的条件下,检验统计量序列 $\{Z_{t_1}, \dots, Z_{t_r}\}$ 渐近服从 Jennison 和 Turnbull^[34]定义的典型联合分布(canonical joint distribution),其中 $0 \leq t_1 \leq t_2 \leq \dots \leq t_r \leq 1$ 。因此,可使用 Jennison 和 Turnbull^[34]专著中的方法来对 DBCD 进行期中分析,从理论上解决了控制犯第一类错误的可能性的问题。

针对罐子模型,Zhu 和 Hu^[37]证明了对两个治疗组的 SEU 模型进行期中分析的理论性质。假设响应变量 ξ_i 的分布参数为 θ 的情况下,SEU 的目标分配比例为 $\rho(\theta)$ 。类似 DBCD 的期中分析,在 t 时刻可通过式(3)得到 θ 的估计 $\hat{\theta}([nt])$,因此可得到此时相应的 $D_{[nt]}(\hat{\theta}([nt]), \xi_{([nt])})$ 和生成矩阵 $H_{[nt]} = H(\hat{\theta}([nt]))$ 。建立假设检验同式(4),并采用检验统计量同式(5),Zhu 和 Hu^[37]证明在一组广

泛适用的条件下,检验统计量序列 $\{Z_{t_1}, \dots, Z_{t_r}\}, 0 \leq t_1 \leq t_2 \leq \dots \leq t_k \leq 1$,渐近服从。Jennison 和 Turnbull^[34]定义的典型联合分布。因此,类似 DBCD, SEU 也可以使用 Jennison 和 Turnbull^[34]中的方法进行期中分析。

两篇文章中的模拟结果显示,DBC D 和 SEU 模型的成组序贯监测都可以控制犯第一类错误的可能性,并同时具有自适应设计和序贯监测各自的优势:(1)提高检验效率,减少总的失败次数;(2)提前终止试验,减少实际样本量。目前关于自适应设计中的序贯监测的研究主要集中在两个治疗组的情况中,在多个试验组的临床试验中如何对自适应设计进行序贯监测还是一个需要进一步研究的重要问题。

2 自适应设计的应用

上文介绍的自适应设计中,一般假设响应变量为二分响应数据和正态响应数据,但是在实际临床试验中,还有很多其他重要的响应数据类型。此外,除了经典的优效试验,非劣效试验也是一类重要的试验类型。由于 DBCD 过程理论性质完善,并具有实现任意给定的目标分配、实际分配比例的方差接近方差下界等优势,近年来,研究者们将 DBCD 拓展到了不同的响应数据类型和试验类型上。本节将介绍 DBCD 在响应变量数据为生存数据^[38-40]、复发事件数据^[41]和有序数据^[42]等类型中的应用,以及非劣效试验中^[43-44]的 DBCD 过程。

2.1 响应变量为生存数据

在肿瘤学、病毒学等领域,很多临床试验的响应变量是生存数据(或时间—事件数据,time-to-event data),如从进入试验起受试者生存的时间或肿瘤发生的时间等。Zhang 和 Rosenberger^[38]最早研究了基于生存数据的自适应设计,针对两个治疗组,在生存数据服从指数分布和 Weibull 分布的情况下,分别提出特定条件下的最优分配比例,并推导出在 DBCD 过程中实际分配比例的渐近性质。随后,Sverdlov 等^[39]和 Sverdlov 等^[40]研究了生存试验中多个试验组的自适应设计。

假设第 $i(i=1, \dots, K)$ 个治疗组中的生存时间

服从参数为 θ_i 的指数分布且具有独立的随机删失机制, $(t_{ij}, \delta_{ij}), j=1, \dots, n_i$, 为该治疗组中的样本, 其中 n_i 为第 i 组的样本量且 $n = \sum_{i=1}^K n_i$. 若第 i 组中第 j 的受试者的数据删失, 则 t_{ij} 为删失时间且 $\delta_{ij} = 0$; 否则 t_{ij} 为生存时间且 $\delta_{ij} = 1$. 记指数分布参数向量 $\theta = (\theta_1, \theta_2, \dots, \theta_k)'$, 样本的似然函数为

$$l(\theta) = \prod_{i=1}^K \prod_{j=1}^{n_i} \left(\theta_i^{-1} \exp\left(-\frac{t_{ij}}{\theta_i}\right) \right)^{\delta_{ij}} \left(\exp\left(-\frac{t_{ij}}{\theta_i}\right) \right)^{1-\delta_{ij}},$$

则 θ_i 的极大似然估计为 $\hat{\theta}_i = \frac{T_i}{\delta_i}$, 其中 $T_i = \sum_{j=1}^{n_i} t_{ij}$ 和 $\delta_i = \sum_{j=1}^{n_i} \delta_{ij}$ 分别为第 i 组中的总观测时间和总事件数。

当 $K=2$ 时, 可建立假设检验 $H_0: \theta_1 = \theta_2$ vs $H_1: \theta_1 \neq \theta_2$, 来比较两个治疗组的治疗效果。Zhang 和 Rosenberger^[38] 使用 Wald 检验统计量

$$Z = \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{\hat{\theta}_1^2}{\delta_1} + \frac{\hat{\theta}_2^2}{\delta_2}}}, \quad (6)$$

进行检验, 并推导相应的最优分配方案。假设 $\varepsilon_i = E(\delta_{ij}), j=1, \dots, n_i$, 则 $E(\delta_i) = n_i E(\delta_{i1}) = n_i \varepsilon_i, i=1, 2$ 。Zhang 和 Rosenberger^[38] 在 $\frac{\theta_1^2}{n\rho\varepsilon_1} + \frac{\theta_2^2}{n(1-\rho)\varepsilon_2} = M$ 的条件下, 提出了三种不同的最优分配方案, 其中 M 为给定的常数, $n\rho = n_1, n(1-\rho) = n_2$,

1. 最小化总危险率, 即最小化 $n\rho\theta_1^{-1} + n(1-\rho)\theta_2^{-1}$, 可得

$$\rho = \frac{\sqrt{\theta_1^3 \varepsilon_2}}{\sqrt{\theta_1^3 \varepsilon_2} + \sqrt{\theta_2^3 \varepsilon_1}},$$

2. 最小化样本量为 n , 可得

$$\rho = \frac{\theta_1 \sqrt{\varepsilon_2}}{\theta_1 \sqrt{\varepsilon_2} + \theta_2 \sqrt{\varepsilon_1}},$$

3. 将数据二分化, 即生存时间小于 c 为失败, 生存时间大于 c 为成功, 最小化 $n_1(1 - \exp(-\frac{c}{\theta_1})) + n_2(1 - \exp(-\frac{c}{\theta_2}))$ ^[45], 可得

$$\rho = \frac{\theta_1 \sqrt{\varepsilon_2(1 - \exp(-\frac{c}{\theta_2}))}}{\theta_1 \sqrt{\varepsilon_2(1 - \exp(-\frac{c}{\theta_2}))} + \theta_2 \sqrt{\varepsilon_2(1 - \exp(-\frac{c}{\theta_1}))}}.$$

类似地, Sverdlov 等^[39] 在 $K \geq 2$ 的情况下, 推导出几类最优分配比例, 包括 D_A -optimal 及基于非线性规划的最优分配等。

若生存时间服从参数为 $(\mu_i, b), i=1, \dots, K$, 的 Weibull 分布, $(t_{ij}, \delta_{ij}), j=1, \dots, n_i$ 的定义类似指数分布模型, 当 $b=1$ 时即为指数分布。令 $z_{ij} = \frac{\log t_{ij} - \mu_i}{b}, \theta = (\mu_1, \mu_2, \dots, \mu_k, b)'$, 则生存数据的似然函数为 $l(\theta) = \prod_{i=1}^K \prod_{j=1}^{n_i} (b^{-1} e^{z_{ij}} \exp(-e^{z_{ij}}))^{\delta_{ij}} (\exp(-e^{z_{ij}}))^{1-\delta_{ij}}$, 参数 θ 的极大似然估计无显式解, 可利用数值方法求解得分方程组 $\frac{\partial \log L(\theta)}{\partial \theta} = 0$ 求出。Zhang 和 Rosenberger^[38] 在 $K=2$ 时, 使用 Wald 检验统计量来检验 μ_1 与 μ_2 是否相等, 并据此推导出两个最优分配方案: 在令检验统计量分母为某个固定常数的条件下, (1) 最小化平均危险率; (2) 最小化平均累计危险率。Sverdlov 等^[40] 在 $K \geq 2$ 情况下, 推导出 D-optimal, 加权最优分配及混合最优分配等方案。

由于必须等待一段时间后才能观测到生存数据(类似于滞后响应), 在分配模型中必须要考虑这一现象对分配过程带来的影响。Zhang 和 Rosenberger^[38] 证明在一组广泛满足的条件下, 一般程度的滞后响应对 DBCD 过程的渐近性质没有影响。同时, 模拟结果^[38-40] 也支持这一结论。

2.2 响应变量为复发事件数据

在药物临床试验中, 复发事件(recurrent event)数据也是一种重要的响应数据类型, 例如哮喘的反复发作、头痛的反复发生及呼吸困难急性发作的反复出现等。在这些临床试验中, 研究目的往往以药物治疗能否有效控制受试者的某临床事件发生频次为最终治疗目的, 因此常用“复发事件”作为临床试验的响应变量。Gao 等^[41] 在两个治疗组的情况下, 考虑响应变量为复发事件数据的 DBCD 过程, 并提出了最优分配方案。

假设 $\{K_{ij}(t), t \geq 0\}$ 为一个计数过程, 记录在 $[0, t]$ 时段内第 $i(i=1, 2)$ 组第 $j(j=1, \dots, n_i)$ 个受试

者的发病次数,其中 n_i 为第 i 个治疗组的样本量且 $n = n_1 + n_2$ 。第 i 个治疗组中的响应变量服从发病率参数为 λ_i 的泊松分布,且 $\lambda_2 = \lambda_1 e^\beta$ 。由于在实践中,复发事件数据的方差往往大于均值,引入均值为 1 的随机效应变量 w_{ij} ,则 $\{K_{ij}(t), t \geq 0\}$ 服从参数为 $\lambda_{ij} = w_{ij} \lambda_1 e^{\beta x_{ij}}$ 的泊松分布,其中 x_{ij} 为分组协变量,即对第 1 组的受试者有 $x_{ij} = 0$,第 2 组的受试者 $x_{ij} = 1$ 。若 w_{ij} 服从均值为 1,方差为 ϕ 的 Gamma 分布,则在给定 x_{ij} 的情况下, $K_{ij}(t)$ 的边际分布为负二项分布

$$P(K_{ij}(t) = k | x_{ij}) = \frac{\Gamma(k + \phi^{-1})}{k! \Gamma(\phi^{-1})} \frac{(\phi \mu_{ij}(t))^k}{(1 + \phi \mu_{ij}(t))^{k + \phi^{-1}}},$$

其中 $\mu_{ij}(t) = \lambda_i t$ 为 t 时刻 $K_{ij}(t)$ 的均值。令 τ_{ij} 表示试验结束时第 i 组第 j 个受试者的观测时长,参数为 $\theta = (\lambda_1, \beta, \phi)'$,则观测数据的对数似然函数为

$$l(\theta) = \sum_{i=1}^2 \sum_{j=1}^{n_i} \left(\ln \frac{\Gamma(k_{ij} + \phi^{-1})}{k_{ij}! \Gamma(\phi^{-1})} + k_{ij} \ln \phi + k_{ij} \right.$$

$$\left. \ln \tau_{ij} + \sum_{i=1}^2 \sum_{j=1}^{n_i} (k_{ij} (\ln \lambda_1 + \beta x_{ij}) - (k_{ij} + \phi^{-1}) \ln(1 + \phi \lambda_1 e^{\beta x_{ij}} \tau_{ij})) \right),$$

其中 k_{ij} 表示第 i 组第 j 个受试者的总发病次数。参数 θ 的极大似然估计需通过数值方法求解。当 $\beta = 0$ 时,两个治疗组的平均发病率相同,故可建立假设检验 $H_0: \beta = 0$ vs $H_1: \beta \neq 0$ 来检验治疗效果是否有差异。Gao 等^[41] 使用 Wald 统计量

$$Z = \frac{\hat{\beta}}{\sqrt{\widehat{\text{Var}}(\hat{\beta})}},$$

进行检验,其中 $\widehat{\text{Var}}(\hat{\beta})$ 为 $\hat{\beta}$ 方差的估计。

在复发事件的临床试验中,受试者在 $[0, \tau_a]$ 时段内依次进入试验,并在接受一段时间治疗后停止试验。在实践中通常有两种停止机制:(1) 等时长治疗:每个受试者在治疗相同长度的时间后停止治疗,即 $\tau_{ij} = \tau_E, i = 1, 2, j = 1, \dots, n_i$;(2) 同时停止:在分配结束后,所有受试者继续治疗一段时间 τ_f 后,同时停止治疗。令 $n_1 = np, n_2 = n(1 - \rho)$,在等时长治疗情况下,

$$\text{Var}(\hat{\beta}) = \frac{V_{E1}}{np} + \frac{V_{E2}}{n(1 - \rho)},$$

其中, $V_{Ei} = \frac{1 + \phi \tau_E \lambda_i}{\tau_E \lambda_i}, i = 1, 2$; 在同时停止治疗的情况下,假设受试者在 $[0, \tau_a]$ 时段内匀速进入试验,在 t 时刻同时停止治疗,则

$$\text{Var}(\hat{\beta}) = \frac{V_{I1}}{np} + \frac{V_{I2}}{n(1 - \rho)},$$

其中, $V_{Ii} = \frac{\phi^2 \lambda_i \tau_a}{\phi \lambda_i \tau_a - \ln \frac{1 + \phi \lambda_i \tau}{1 + \phi \lambda_i \tau_f}}, i = 1, 2, \tau_f = \tau - \tau_a$ 。

考虑两种优化问题:

1. 最大化检验的功效,即在样本量固定的情况下,最小化 $\text{Var}(\hat{\beta})$, 可得

$$\rho = \frac{\sqrt{V_1}}{\sqrt{V_1} + \sqrt{V_2}},$$

2. 最小化总发病次数,即在固定 $\text{Var}(\hat{\beta})$ 的情况下,最小化 $n_1 \lambda_1 + n_2 \lambda_2$, 可得

$$\rho = \frac{\sqrt{\frac{V_1}{\lambda_1}}}{\sqrt{\frac{V_1}{\lambda_1}} + \sqrt{\frac{V_2}{\lambda_2}}},$$

将 V_{E1}, V_{E2} 和 V_{I1}, V_{I2} 分别代入分配方案的 V_1, V_2 中,即可得到相应条件下的最优分配方案。模拟结果表明,与均衡分配相比两种分配都能够分配较多的受试者到较好的治疗组中。此外,第一个最优分配方案能够提高检验的功效,而第二个最优分配方案能够在不损失太多检验功效的情况下,可以分配比第一个最优分配方案更多的受试者到较好的治疗组中。

2.3 响应变量为有序数据

在临床试验中,有序数据也是一种常见的响应数据类型,如通过调查问卷形式对受试者的抑郁程度进行测量,可能得到的测量结果有:轻度抑郁,中度抑郁,中重度抑郁及重度抑郁。Lu 等^[42] 采用 Weibull 潜变量模型(latent Weibull model)来比较治疗组的治疗效果差异,并提出相应的最优分配,通过 DBCD 过程来分配受试者。

假设共有两个治疗组, Y_i 表示第 i 个治疗组的响应变量,可能的取值为 $k, k = 1, \dots, K$,假设 k 越大表示治疗结果越好。在潜变量模型中,通常假设潜

在的真实响应变量 Z_i 是一个连续变量, 而实际观测到的有序响应数据 Y_i 是真实响应变量的一种表现形式, 二者之间的关系为 $Y_i = k$ 当且仅当 $\alpha_{k-1} < Z_i \leq \alpha_k, k = 1, \dots, K, i = 1, 2$, 其中间断点 $-\infty = \alpha_0 < \alpha_1 < \dots < \alpha_{k-1} < \alpha_k = \infty$ 在两个治疗组中的取值是一样的。若 Z_i 服从参数为 $\omega_i = (\eta_i, \beta_i)', i = 1, 2$, 的 Weibull 分布, $G(z | \omega_i) = 1 - \exp\left(-\left(\frac{z}{\eta_i}\right)^{\beta_i}\right), z > 0$, n_{ik} 表示第 i 组中响应变量取值为 k 的受试者数目, 则对数似函数为

$$l(\theta) = \sum_{i=1}^2 \sum_{k=1}^K n_{ik} \log(\pi_{ik}(\theta)),$$

其中, $\theta = (\alpha', \omega'_1, \omega'_2)', \alpha = (\alpha_1, \dots, \alpha_{K-1})', \pi_{ik}(\theta)$ 为第 i 组中响应变量取值为 k 的概率, 即 $\pi_{ik}(\theta) = G(\alpha_k | \omega_i) - G(\alpha_{k-1} | \omega_i)$ 。记 $\hat{\theta} = (\hat{\alpha}', \hat{\omega}'_1, \hat{\omega}'_2)'$ 为参数向量 θ 的估计, 可得

$$n^{1/2}(\hat{\omega}_i - \omega_i) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Omega}_i), \mathbf{\Omega}_i = \begin{pmatrix} a_{1,i} & a_{2,i} \\ a_{2,i} & a_{3,i} \end{pmatrix},$$

其中 a_{ij} 的形式较复杂, 具体表达式参见 Lu 等^[42] 的附录。

由于 Weibull 分布非对称, 可通过比较分布的中位数来比较两个治疗组的治疗效果。令 M_i 表示第 i 组潜在 Weibull 分布的中位数, 则 $M_i = \eta_i (\ln 2)^{1/\beta_i}$, 且有

$$n^{1/2}(\hat{M}_i - M_i) \xrightarrow{D} N(0, \gamma_i^2),$$

其中, $\gamma_i^2 = (\ln 2)^{2/\beta_i} a_{1,i} - 2 \frac{\eta_i}{\beta_i^2} (\ln 2)^{2/\beta_i} (\ln \ln 2) a_{2,i} + \eta_i^2 / \beta_i^4 (\ln 2)^{2/\beta_i} (\ln \ln 2)^2 a_{3,i}$ 。建立假设检验 $H_0: M_1 = M_2$ vs $H_1: M_1 \neq M_2$ 及检验统计量

$$Z = \frac{\hat{M}_2 - \hat{M}_1}{\sqrt{\frac{\hat{\gamma}_1^2}{n_1} + \frac{\hat{\gamma}_2^2}{n_2}}}$$

令 $n_1 = n\rho, n_2 = n(1 - \rho)$, 考虑三种最优分配比例

1. 最大化检验的功效, 可得

$$\rho = \frac{\gamma_1}{\gamma_1 + \gamma_2},$$

2. 最小化接受较差治疗的人数, 即固定方差的

情况下, 最小化 $\frac{n_1}{M_1} + \frac{n_2}{M_2}$,

$$\rho = \frac{\sqrt{M_1} \gamma_1}{\sqrt{M_1} \gamma_1 + \sqrt{M_2} \gamma_2},$$

3. 修正的第二种最优分配, 即固定方差的情况

下, 最小化 $\frac{n_1}{M_1/\gamma_1} + \frac{n_2}{M_2/\gamma_2}$,

$$\rho = \frac{\sqrt{M_1} \gamma_1}{\sqrt{M_1} \gamma_1 + \sqrt{M_2} \gamma_2}.$$

模拟结果显示, 新的最优分配与均衡分配相比, 能够在检验功效基本相同的情况下分配更多的受试者接受好的治疗。其中第三种最优分配在不同的参数假设下均表现良好, 具有较强的稳健性。

2.4 非劣效试验

非劣效临床试验 (non-inferiority clinical trials, NI) 的目的是, 在标准治疗方案已经存在的情况下, 验证一种新的治疗方案的治疗效果与标准治疗的效果接近。这种新治疗方案的疗效与标准方案相差不多, 但在其它方面应具有明显的优势, 比如具有更少的副作用, 或成本更低, 或更容易实施等。因此, 非劣效临床试验得到了越来越多的关注。

Xu 等^[44] 假设一个 NI 试验中包含 K 个新治疗组 ($E_k, k = 1, \dots, K$), 一个标准治疗组 (R) 和一个安慰剂组 (P), 第 i 组第 j 个受试者的响应变量为 $X_{ij} = \mu_i + \varepsilon_{ij}, i = E_1, \dots, E_k, R, P, j = 1, \dots, n_i$, 其中, μ_i 为第 i 组的均值, n_i 为第 i 组的样本量, 随机变量 ε_{ij} 服从均值为 0, 方差为 σ_i^2 的正态分布。不失一般性, 假设 μ_i 越大表示治疗效果越好。记 \bar{X}_i 为第 i 组的样本均值, S_i^2 为第 i 组样本方差, 是 σ_i^2 的无偏估计且与 \bar{X}_i 独立。在 NI 试验中, 需要同时建立 $K + 1$ 个假设检验

$$H_{0k}: \mu_{E_k} \leq \mu_R - M_2, \quad k = 1, \dots, K,$$

$$H'_0: \mu_R \leq \mu_P + M_1,$$

vs

$$H_{1k}: \mu_{E_k} > \mu_R - M_2, \quad k = 1, \dots, K,$$

$$H'_1: \mu_R > \mu_P + M_1,$$

其中 M_1, M_2 为已知常数。检验统计量为

$$T_{E_k} = \frac{\bar{X}_{E_k} - \bar{X}_R + M_2}{\sqrt{\frac{S_{E_k}^2}{n_{E_k}} + \frac{S_R^2}{n_R}}}, \quad k = 1, \dots, K,$$

$$T_P = \frac{\bar{X}_R - \bar{X}_P - M_1}{\sqrt{\frac{S_P^2}{n_P} + \frac{S_R^2}{n_R}}},$$

其中, $T_i, i = E_1, \dots, E_k, R, P$, 服从 t 分布。Xu 等^[43] 定义总方差

$$\text{TotVar} = \sum_{k=1}^K \text{Var}(\hat{\mu}_{E_k} - \hat{\mu}_R + M_2) + \text{Var}(\hat{\mu}_R - \hat{\mu}_P - M_1) = \frac{\sigma_{E_1}^2}{n_{E_1}} + \dots + \frac{\sigma_{E_G}^2}{n_{E_G}} + (G+1) \frac{\sigma_R^2}{n_R} + \frac{\sigma_P^2}{n_P},$$

并提出了一种最优分配, XHC 分配: 在固定 TotVar 的情况下, 分配更多的受试者到治疗效果较好的治疗组中, 即最小化 $\sum_{k=1}^K \frac{n_{E_k}}{e^{qE_k}} + \frac{n_P}{e^{qP}} + \frac{n_R}{e^{qR}}$, 可得

$$\rho_P = \frac{\sigma_P \sqrt{e^{qP}}}{\sum_{j=1}^G \sigma_{E_j} \sqrt{e^{qE_j}} + \sigma_P \sqrt{e^{qP}} + \sigma_R \sqrt{(G+1)e^{qR}}},$$

$$\rho_{E_k} = (1 - \rho_P) \frac{\sigma_{E_k} \sqrt{e^{qE_k}}}{\sum_{j=1}^G \sigma_{E_j} \sqrt{e^{qE_j}} + \sigma_P \sqrt{e^{qP}} + \sigma_R \sqrt{(G+1)e^{qR}}}, \quad k = 1, \dots, G,$$

$$\rho_R = (1 - \rho_P) \frac{\sigma_R \sqrt{(G+1)e^{qR}}}{\sum_{j=1}^G \sigma_{E_j} \sqrt{e^{qE_j}} + \sigma_P \sqrt{e^{qP}} + \sigma_R \sqrt{(G+1)e^{qR}}},$$

其中, q 为调节参数, 当 q 较大时, 会分配较多受试者进入标准治疗组, Xu 等^[43] 建议 q 的取值为 M_2 的函数 $q = \frac{1}{2M_2}$ 。在 TotVal 中标准治疗组的方差权重较大, 为此 Xu 等^[44] 定义了新的总方差 NTotVal,

$$\text{NTotVar} = \sum_{k=1}^K \text{Var}(\hat{u}_{E_k}) + \text{Var}(\hat{u}_R) + \text{Var}(\hat{u}_P) = \frac{\sigma_{E_1}^2}{n_{E_1}} + \dots + \frac{\sigma_{E_k}^2}{n_{E_k}} + \frac{\sigma_R^2}{n_R} + \frac{\sigma_P^2}{n_P},$$

并在固定 NTotVar 为某个常数的条件下, 最小化 $\sum_{k=1}^K \frac{n_{E_k}}{e^{u_{E_k}/\sigma_{E_k}}} + \frac{n_P}{e^{u_P/\sigma_P}} + \frac{n_R}{e^{u_R/\sigma_R}}$, 得到 MXHC (modified XHC) 最优分配比例

$$\rho_P = \frac{\sigma_P \sqrt{e^{u_P/\sigma_P}}}{\sum_{j=1}^G \sigma_{E_j} \sqrt{e^{u_{E_j}/\sigma_j}} + \sigma_P \sqrt{e^{u_P/\sigma_P}} + \sigma_R \sqrt{e^{u_R/\sigma_R}}},$$

$$\rho_{E_k} = (1 - \rho_P) \frac{\sigma_{E_k} \sqrt{e^{u_{E_k}/\sigma_{E_k}}}}{\sum_{j=1}^G \sigma_{E_j} \sqrt{e^{u_{E_j}/\sigma_R}} + \sigma_R \sqrt{e^{u_R/\sigma_R}}}, \quad k = 1, \dots, K,$$

$$\rho_R = (1 - \rho_P) \frac{\sigma_R \sqrt{e^{u_R/\sigma_R}}}{\sum_{j=1}^G \sigma_{E_j} \sqrt{e^{u_{E_j}/\sigma_R}} + \sigma_R \sqrt{e^{u_R/\sigma_R}}},$$

从模拟结果中可以看到, 当标准治疗组的治疗效果最好时, 相比于偏均衡分配, XHC 和 MXHC 分配方案都能提高检验效率并减少分配到治疗效果较差的受试者数目; 在同方差情况下, XHC 要优于 MXHC 分配方案, 但是由于 MXHC 的稳健性更强, 在没有方差的先验信息的情况下, MXHC 分配方案是一个比较好的选择。值得注意的是, 在非劣效试验的分配方案中提出的最优分配, 是在固定“总方差”的情况下得到的。这里两种总方差的定义的都是主观的, 固定总方差并不等价于固定检验功效。因此, 如何在固定检验功效的情况下, 得到相应的最优分配还需要进一步的研究。

3 总结及展望

3.1 总结

由于自适应设计在伦理道德方面等的优势, 自提出起就受到学术界和工业界的广泛关注。早期的罐子模型简单直观, 但由于模型本身的限制, 无法完全解决实际应用遇到的问题。此后, 学者们不断改进自适应设计, 完善自适应设计的理论框架, 并进一步将其推广到不同响应数据类型及试验类型中。

近二十年来, 自适应设计的目标分配比例不仅考虑将更多的受试者分配到好的治疗组, 还要考虑检验功效等方面的问题。在目标驱动的罐子模型中, SEU 通过调整添加球的数目, 能够使实际分配比例达到目标比例, 随后 Zhang 等^[31] 提出的 IMU 模型包含了 SEU 在内的几乎所有常见罐子模型, 为罐子模型建立了一个系统的理论框架。另外, Hu 和 Rosenberger^[17] 发现实际分配比例的方差越小检验功效越大, Hu 等^[28] 给出了实际分配比例的方差的下界, 因此在目标分配比例相同的情况下, 可通过实际分配比例方差来评价各模型的优劣。

目前研究最广泛的自适应设计是基于序贯估计的 DBCD 过程,它能使实际分配比例达到给定的目标比例,且实际分配比例的方差接近方差下界。同时,BCD 过程在广泛适用的条件下具有优良的渐近性质,能够在控制第一类错误的条件下进行序贯监测,且一般程度的滞后响应和时间趋势异质性对其渐近性质无影响。此外,不同于罐子模型主要适用于二分响应数据,BCD 过程还适用于各种不同响应数据类型,如连续型数据、生存数据、复发事件数据及有序数据等。在理论和应用方面,以 DBCD 过程为主自适应设计已取得了丰硕的成果,但也存在一些未解决的问题,未来还需要研究的问题有下面几个方面。

3.2 展望

3.2.1 贝叶斯自适应设计

本文介绍的自适应设计主要是基于频率学派的思想,在临床试验中,贝叶斯适应性设计(Bayesian adaptive designs)也是一类重要的临床试验设计。有关贝叶斯框架下的适应性设计,可参见 Tompson^[10]、Atkinson 和 Biswas^[46]、Thall 和 Wathen^[47]、Huang 等^[48]及 Yuan 等^[49]等文献,此外, Berry 等^[50]在专著中详细总结了贝叶斯适应性设计在各阶段临床试验中的应用。贝叶斯方法可以根据先验概率和已知数据获得到后验概率,并根据后验概率调整试验方案,因此在小样本情况下,可得到具有较高精度的统计推断。由于在节约时间和成本方面具有较大的优势,贝叶斯方法得到了药物研发领域的广泛关注^[51],并被越来越多地应用到 I 期和 II 期临床试验中。其中, Giles 等^[52]和 Zhou 等^[53]在临床试验中采用了贝叶斯适应性设计,其试验结果均显示贝叶自适应设计能够快速识别出有效的治疗方案,并将更多的受试者分配到较好的试验组中。

如何将贝叶斯思想和自适应设计结合,提出一类能够结合两者优势的试验设计,是一个有意义的问题, Xiao 等^[54]和 Kim 等^[55]已在这方面进行了初步的工作。Xiao 等^[54]将 Thall 和 Wathen^[47]中基于

后验竞争概率 (posterior competing probability) $\hat{p}_{i-1} = P(\theta_1 > \theta_2 | \xi_1, \dots, \xi_{i-1})$ 的分配概率 $\hat{p}_{i-1} = \frac{\hat{p}_{i-1}^\tau}{\hat{p}_{i-1}^\tau + (1 - \hat{p}_{i-1})^\tau}$ 运用到 DBCD 过程当中,使得能够在控制犯第一类错误的可能性的情况下,分配更多的受试者到较好的治疗组中,并且实际分配比例具有较小的方差。另一方面, Kim 等^[55]在贝叶斯自适应设计框架下引入 DBCD 过程,使得该设计既可以适应性地运用历史对照数据,又可以在试验组有效的情况下将更多的受试者分配到其中。这些工作体现了将贝叶斯思想和 DBCD 等自适应设计结合是一个很有意义的方向,然而目前这类工作都存在一些不足。Xiao 等^[54]提出的贝叶斯自适应设计的实际分配比例收敛到一个事先给定的分配界限,而不是由参数决定的最优分配。Kim 等^[55]的设计则需要高质量的历史对照数据来提高检验效率和降低样本量。如何改进这些贝叶斯自适应设计还是一个需要进一步研究的问题。此外,在 Xiao 等^[54]和 Kim 等^[55]提出的设计中,响应变量都是二分的,在其他数据类型的响应变量情况下,如何将贝叶斯思想和自适应设计结合还有待更深入的探讨。

3.2.2 协变量调整—自适应设计(CARA)

传统的医疗模式往往注重于大样本人群的标准治疗,较少考虑到病人间的个体差异。而精准医疗 (precision medicine) 要求在药物研发过程中考虑病人的差异化信息,在检验药物有效性的同时,指导个性化医疗药物的使用方案。因此,随着精准医疗的不断发展,越来越多的临床试验开始关注协变量(或生物标记)与治疗之间的关系,一类新的适应性随机化方法——协变量调整—自适应设计 (CARA) 得到了广泛的关注。在 CARA 设计中,分配新的受试者的概率不仅与之前受试者的响应数据有关,还与当前受试者和之前受试者的协变量有关。此类设计的主要目的是,在协变量与治疗结果相关的情况下,根据受试者的协变量,以较大的概率将其分配到治疗效果更好治疗组中,因此从伦理道德方面来看, CARA 能够照顾到每个受试者的最大利益。

在已有的几类 CARA 设计中,Zhang 等^[56]提出的模型是根据给定的目标分配函数来决定每个受试者的分配概率,其中目标分配函数是响应变量和协变量的函数,该设计的实际分配比例的方差偏大。随后,Zhang 和 Hu^[57]在 DBCD 设计中引入协变量,提出了 CADBCD 过程,有效减小了实际分配比例的方差。这一设计具有完整的理论性质,但是在实际应用中还存在一些问题,例如在试验初期受试者的协变量未知时,应如何计算最优分配比例等。Zhang 等^[56]建议先计算在没有协变量情况下的最优分配估计,然后通过协变量调整来得到相应的最优分配。Hu 等^[58]提出了两个度量检验功效和伦理道德的一般标准,在此基础上得到目标分配比例,建立了一族一般的 CARA 模型,并在协变量为分类变量的情况下,推导出 CARA 模型的渐近性质。在协变量为其他数据类型时,如何建立相应的 CARA 模型还需要做进一步研究。

此外,协变量信息可以根据其在临床试验中的不同作用分为两类:预后(prognostic)协变量的作用是通过平衡协变量使得不同的治疗组更加相似而增加试验的可信度;预测(predictive)协变量的作用是基于有效性和伦理性为受试者选择更合适的治疗组。目前已有文献中的 CARA 试验设计主要考虑运用预测协变量的信息优化试验组选择,而忽视了预后协变量的均衡问题。因此如何在自适应试验当中同时考虑两类协变量以及响应变量信息,即在平衡预后协变量的同时分配更多的受试者到效果更好的治疗组中,在今后的研究中值得我们关注。

3.2.3 序贯监测

临床试验中的期中分析和序贯监测具有重要的理论和实践意义,Zhu 和 Hu^[35,37]证明了在自适应设计中,检验统计量序列 $\{Z_{t_1}, \dots, Z_{t_r}\}, 0 \leq t_1 \leq t_2 \leq \dots \leq t_r \leq 1$ 在一组广泛适用的条件下,渐近服从 Jennison 和 Turnbull^[34]定义的典型联合分布,从而从理论上解决了控制自适应设计序贯监测中第一类错误的可能性,及其他有关检验统计量序列分布的理论问题。但这些理论证明都是基于响应变量为二分数

据或连续数据及参数的矩估计,未考虑其他类型的数据或参数估计形式。一个有意义的问题是,当响应数据为生存数据等数据类型时,检验统计量序列的分布形式是否还是典型联合分布。另一方面,Hampson 和 Jennison^[59]发现存在滞后响应时,序贯监测在减少样本量等方面的优势会减低。在目前 DBCD 序贯监测^[35,37]的研究中,均未考虑滞后响应对序贯监测的影响,因此在滞后响应存在的情况下,自适应设计中的序贯监测是否有效是一个需要重视问题。此外,现有的自适应设计的期中分析主要集中在两个治疗组的试验中。在非劣效试验或多重比较试验中,通常涉及多个检验统计量序列,这就增加了期中分析的复杂性。因此,如何对多个治疗组的试验进行期中分析,决定是否继续试验或是否减少试验组,也都是自适应设计中亟需解决的问题。

3.2.4 缺失数据

在临床试验中,经常会遇到缺失数据的情况。在自适应设计中,若响应变量数据缺失,则可能出现参数估计有偏,实际分配比例偏离目标分配比例等问题,进而影响自适应设计在伦理道德方面的优势及检验效率。Molenberghs 和 Kenward^[60]在书中详细介绍了临床试验中各种处理缺失数据方法的应用,但是他们主要关注的是固定设计(fixed design)。与固定设计相比,自适应设计处理缺失数据的难点在于,在随机化分配过程中需要序贯地分析缺失数据,并且缺失数据的影响会在分配过程中不断累积,对最终的统计分析造成困难。找出数据缺失的原因能够帮助分析已有的数据,若数据随机缺失,则已知数据仍具有代表性,但是如果数据缺失具有系统性,则分析过程将会很复杂。Ma^[61]在完全随机缺失和随机缺失机制下,分析了缺失响应数据对 DBCD 过程的影响:在完全随机缺失机制下,BCD 过程稳健但波动性增大;在随机缺失机制下,参数估计有偏,需用基于极大似然的方法对缺失数据进行序贯分析。在随机化试验设计中,研究缺失数据的影响并找到合适的解决方案对实际应用有着深远的意义。

参考文献:

- [1] Chow S C, Chang M. Adaptive Design Methods in Clinical Trials (2nd ed.) [M]. Boca Raton, FL: CRC Press, 2012.
- [2] Bothwell L E, Avorn J, Khan N F, et al. Adaptive design clinical trials; A review of the literature and Clinical-Trials. gov[EJ]. BMJ Open, 2018, 8(2): e018320.
- [3] 颜虹, 夏结来, 于莉莉. 临床试验中适应性设计研究进展[J]. 中华预防医学杂志, 2008, 42(S): 16-25.
- [4] 王维亭, 郝春华, 汤立达. 临床适应性设计及在新药研发中的应[J]. 现代药物与临床, 2010, 25(5): 334-339.
- [5] 杜艾桦. 药物临床试验适应性设计的科学性与伦理性问题[J]. 药物流行病学杂志, 2017, 26(12): 837-840.
- [6] Rosenberger W F, Lachin J M. Theory and Practice: Randomization in Clinical Trials[M]. Hoboken, NJ: John Wiley & Sons, 2015.
- [7] Hu F, Rosenberger W F. The Theory of Response - Adaptive Randomization in Clinical Trials[M]. Hoboken, NJ: John Wiley & Sons, 2006.
- [8] 于莉莉, 薛富波, 王素珍, 等. 临床试验适应性设计的反应变量—适应性随机化方法简介[J]. 中国卫生统计, 2008, 25(5): 534-536.
- [9] 刘晓燕, 陈峰, 魏永越, 等. 响应—自适应随机化分组方法[J]. 中国临床药理学与治疗学, 2008, 13(6): 684-689.
- [10] Thompson W R. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples[J]. Biometrika, 1933, 25(3/4): 285-294.
- [11] Robbins H. Some aspects of the sequential design of experiments[J]. Bulletin of the American Mathematical Society, 1952, 58(5): 527-535.
- [12] Anscombe F J. Sequential medical trials[J]. Journal of the American Statistical Association, 1963, 58(302): 365-383.
- [13] Colton T. A model for selecting one of two medical treatments[J]. Journal of the American Statistical Association, 1963, 58(302): 388-400.
- [14] Zelen M. Play the winner rule and the controlled clinical trial[J]. Journal of the American Statistical Association, 1969, 64(325): 131-146.
- [15] Wei L J, Durham S. The randomized play - the - winner rule in medical trials[J]. Journal of the American Statistical Association, 1978, 73(364): 840-843.
- [16] Ivanova A. A play - the - winner - type urn design with reduced variability[J]. Metrika, 2003, 58(1): 1-13.
- [17] Hu F, Rosenberger W F. Optimality, variability, power[J]. Journal of the American Statistical Association, 2003, 98(463): 671-678.
- [18] Athreya K B, Karlin S. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems[J]. The Annals of Mathematical Statistics, 1968, 39(6): 1801-1817.
- [19] Bai Z D, Hu F. Asymptotics in randomized urn models[J]. The Annals of Applied Probability, 2005, 15(1B): 914-940.
- [20] Ivanova A, Rosenberger W F, Durham S D, et al. A birth and death urn for randomized clinical trials; Asymptotic methods[J]. The Indian Journal of Statistics (Series B), 2000, 62(1): 104-118.
- [21] Zhang L X, Chan W S, Cheung S H, et al. A generalized drop - the - loser urn for clinical trials with delayed responses[J]. Statistica Sinica, 2007, 17(1): 387-409.
- [22] Melfi V F, Page C. Variability in adaptive designs for estimation of success probabilities[A]. New Developments and Applications in Experimental Design[C]. Hayward, CA: Institute of Mathematical Statistics, 1998: 106-114.
- [23] Rosenberger W F, Stallard N, Ivanova A, et al. Optimal adaptive designs for binary response trials[J]. Biometrics, 2001, 57(3): 909-913.
- [24] Tymofeyev Y, Rosenberger W F, Hu F. Implementing optimal allocation in sequential binary response experiments[J]. Journal of the American Statistical Association, 2007, 102(477): 224-234.
- [25] Eisele J R. The doubly adaptive biased coin design for sequential clinical trials[J]. Journal of Statistical Planning and Inference, 1994, 38(2): 249-261.
- [26] Eisele J R, Woodroffe M B. Central limit theorems

for doubly adaptive biased coin designs[J]. *The Annals of Statistics*, 1995, 23(1): 234 – 254.

[27] Hu F, Zhang L X. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials[J]. *The Annals of Statistics*, 2004, 32(1): 268 – 301.

[28] Hu F, Rosenberger W F, Zhang L X. Asymptotically best response – adaptive randomization procedures[J]. *Journal of Statistical Planning and Inference*, 2006, 136(6): 1911 – 1922.

[29] Hu F, Zhang L X, He X. Efficient randomized – adaptive designs[J]. *The Annals of Statistics*, 2009, 37(5A): 2543 – 2560.

[30] Zhang L X, Hu F, Cheung S H. Asymptotic theorems of sequential estimation – adjusted urn models[J]. *The Annals of Applied Probability*, 2006, 16(1): 340 – 369.

[31] Zhang L X, Hu F, Cheung S H, et al. Immigrated urn models – theoretical properties and applications[J]. *The Annals of Statistics*, 2011, 39(1): 643 – 671.

[32] Melfi V F, Page C, Geraldes M. An adaptive randomized design with application to estimation[J]. *Canadian Journal of Statistics*, 2001, 29(1): 107 – 116.

[33] 肖意可, 胡飞芳, 刘中强, 等. 适应性设计在新药临床试验中的应[J]. *中国新药杂志*, 2017, 26(4): 389 – 393.

[34] Jennison C, Turnbull B W. *Group Sequential Methods with Applications to Clinical Trials*[M]. Boca Raton, FL: Chapman & Hall/CRC, 1999.

[35] Zhu H, Hu F. Sequential monitoring of response – adaptive randomized clinical trials[J]. *The Annals of Statistics*, 2010, 38(4): 2218 – 2241.

[36] Gwise T, Hu F, Hu J. Optimal biased coins for two – arm clinical trials[J]. *Statistics and Its Interface*, 2008, 1(1): 125 – 135.

[37] Zhu H, Hu F. Interim analysis of clinical trials based on urn models[J]. *Canadian Journal of Statistics*, 2012, 40(3): 550 – 568.

[38] Zhang L, Rosenberger W F. Response – adaptive randomization for survival trials: The parametric approach[J]. *Journal of the Royal Statistical Society (Series C): Applied Statistics*, 2007, 56(2): 153 – 165.

[39] Sverdlov O, Tymofeyev Y, Wong W K. Optimal response – adaptive randomized designs for multiarmed survival trials[J]. *Statistics in Medicine*, 2011, 30(24): 2890 – 2910.

[40] Sverdlov O, Ryzhnik Y, Wong W K. Efficient and ethical response – adaptive randomization designs for multi – arm clinical trials with Weibull time – to – event outcomes[J]. *Journal of Biopharmaceutical Statistics*, 2014, 24(4): 732 – 754.

[41] Gao J, Su P F, Hu F, et al. Adaptive treatment allocation for comparative clinical studies with recurrent events data[J]. *Biometrics*, 2020, 76(1): 183 – 196.

[42] Lu T Y, Chung K P, Poon W Y, et al. Response – adaptive treatment allocation for clinical studies with ordinal responses[J]. *Statistical Methods in Medical Research*, 2020, 29(2): 359 – 373.

[43] Xu W, Hu F, Cheung S H. Adaptive designs for non – inferiority trials with multiple experimental treatments[J]. *Statistical Methods in Medical Research*, 2018, 27(11): 3255 – 3270.

[44] Xu W, Gao J, Hu F, et al. Response – adaptive treatment allocation for non – inferiority trials with heterogeneous variances[J]. *Computational Statistics & Data Analysis*, 2018, 124: 168 – 179.

[45] Biswas A, Mandal S. Optimal adaptive designs in phase III clinical trials for continuous responses with covariates[A]. In: Di Bucchianico A W H, Läuter H. *moDa 7 – Advances in Model – oriented Design and Analysis*[C]. Physica, Heidelberg: Springer, 2004: 51 – 59.

[46] Atkinson A C, Biswas A. Bayesian adaptive biased – coin designs for clinical trials with normal responses[J]. *Biometrics*, 2005, 61(1): 118 – 125.

[47] Thall P F, Wathen J K. Practical Bayesian adaptive randomisation in clinical trials[J]. *European Journal of Cancer*, 2007, 43(5): 859 – 866.

[48] Huang X, Ning J, Li Y, et al. Using short – term response information to facilitate adaptive randomization for survival clinical trials[J]. *Statistics in Medicine*, 2009, 28(12): 1680 – 1689.

[49] Yuan Y, Huang X, Liu S. A Bayesian response –

adaptive covariate – balanced randomization design with application to a leukemia clinical trial[J]. *Statistics in Medicine*, 2011, 30(11): 1218 – 1229.

[50]Berry S M, Carlin B P, Lee J J, et al. *Bayesian Adaptive Methods for Clinical Trials* [M]. Boca Raton, FL: CRC Press, 2010.

[51]Woodcock J. FDA introductory comments; Clinical studies design and evaluation issues [J]. *Clinical Trials*, 2016, 2(4): 273 – 275.

[52]Giles F J, Kantarjian H M, Cortes J E, et al. Adaptive randomized study of idarubicin and cytarabine versus troxacitabine and cytarabine versus troxacitabine and idarubicin in untreated patients 50 years or older with adverse karyotype acute myeloid leukemia[J]. *Journal of Clinical Oncology*, 2003, 21(9): 1722 – 1727.

[53]Zhou X, Liu S, Kim E S, et al. Bayesian adaptive design for targeted therapy development in lung cancer; A step toward personalized medicine [J]. *Clinical Trials*, 2008, 5(3): 181 – 193.

[54]Xiao Y, Liu Z, Hu F. Bayesian doubly adaptive randomization in clinical trials[J]. *Science China Mathematics*, 2017, 60(12): 2503 – 2514.

[55]Kim M O, Harun N, Liu C, et al. Bayesian selec-

tive response – adaptive design using the historical control[J]. *Statistics in Medicine*, 2018, 37(26): 3709 – 3722.

[56]Zhang L X, Hu F, Cheung S H, et al. Asymptotic properties of covariate – adjusted response – adaptive designs[J]. *The Annals of Statistics*, 2007, 35(3): 1166 – 1182.

[57]Zhang L X, Hu F. A new family of covariate – adjusted response adaptive designs and their properties[J]. *Applied Mathematics – A Journal of Chinese Universities*, 2009, 24(1): 1 – 13.

[58]Hu J, Zhu H, Hu F. A unified family of covariate – adjusted response – adaptive designs based on efficiency and ethics[J]. *Journal of the American Statistical Association*, 2015, 110(509): 357 – 367.

[59]Hampson L V, Jennison C. Group sequential tests for delayed responses[J]. *Journal of the Royal Statistical Society (Series B): Statistical Methodology*, 2012, 75(1): 3 – 54.

[60]Molenberghs G, Kenward M G. *Statistics in Practice: Missing Data in Clinical Studies* [M]. Chichester, UK: John Wiley & Sons, 2007.

[61]Ma Z. *Missing data and adaptive designs in clinical studies* [D]. Charlottesville, VA: University of Virginia, 2013.

Recent Progress and Prospect of Response – Adaptive Randomization in Clinical Trials

Gao Jingya Ma Wei Zhang Shaohong Hu Feifang

Abstract: Because of its three advantages: Ethical, efficient and flexible, response – adaptive randomization (RAR) has recently attracted a great deal of attention in clinical research. In this paper, we first introduce two main types of RAR, namely, randomized urn models and the sequential – estimation model. Then we focus on the recent advances of RAR in the literature: (i) sequential monitoring of clinical trials based on RAR; and (ii) the applications of RAR for survival data, recurrent event data, and ordinal data and in non – inferiority trials. Lastly, we discuss a few unsolved issues and suggest some future research directions.

Key words: response – adaptive randomization; randomized urn model; doubly – adaptive biased coin design; optimal allocation