

# 适应性富集设计中的效应点估计方法比较\*

南京医科大学公共卫生学院生物统计学系(211166) 王爽<sup>#</sup> 仲子航<sup>#</sup> 朱子涵 焦阳 柏建岭<sup>△</sup> 于浩<sup>△</sup>

**【摘要】目的** 探讨适应性富集设计中的效应点估计方法,为富集设计下的疗效估计提供参考。**方法** 通过模拟适应性富集设计中的疗效数据,比较极大似然方法、第二阶段估计方法、Lindley 估计方法、Bootstrap 参数方法和条件无偏估计方法这五种方法在估计药物疗效方面的统计学性能。**结果** 极大似然估计方法往往产生较大偏倚,且容易受参数影响;Lindley 估计方法和 Bootstrap 参数方法偏倚较小,但 Lindley 估计方法可能低估药物疗效;第二阶段估计方法和条件无偏估计方法的估计几乎无偏。除了第二阶段估计方法外,其他方法的均方误差均较小。**结论** 在适应性富集设计中,条件无偏估计方法的偏倚较小,均方根误差在可接受范围内,因此推荐使用该方法来估计药物疗效。

**【关键词】** 富集设计 效应估计 期中分析 亚组

**【中图分类号】** R195.1 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2025.01.007

## A Comparison of Point Estimation Methods in Adaptive Enrichment Designs

Wang Shuang, Zhong Zihang, Zhu Zihan, et al (Department of Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166)

**【Abstract】Objective** To explore methods for point estimation in adaptive enrichment design, providing a reference for effect estimation in this context. **Methods** We simulated efficacy data under adaptive enrichment design and subsequently estimated the effects using five methods: maximum likelihood method, stage 2 method, Lindley's method, parametric Bootstrap method, and the uniformly minimum variance conditionally unbiased estimated (UMVCUE) method. The performance of these methods was further evaluated. **Results** The simulation study showed that the maximum likelihood method has a large bias and is very sensitive to some parameters, while Lindley's method tends to underestimate and the parametric bootstrap method sometimes overestimates. The stage 2 method and UMVCUE are almost unbiased. In terms of mean squared error (MSE), all methods perform well except stage 2 estimator. **Conclusion** In adaptive enrichment design, the UMVCUE performs well in terms of bias and MSE. Therefore, we recommend it as the point estimator for the treatment effect in adaptive enrichment design.

**【Key words】** Enrichment designs; Point estimation; Interim analysis; Subgroup

随着医学分子生物学的快速发展,肿瘤治疗正从传统的特定瘤种化疗逐渐转向针对特定生物标志物的靶向治疗。以“广谱治疗”为特点的药物临床试验设计思路已不能适应当前以分子靶向和基因驱动为特征的新药研发趋势。针对这种情况,在临床试验中提出了适应性富集设计。该设计能够根据期中分析的结果,按照预先设定的标准对目标人群进行适应性调整,从而决定试验后续阶段的目标人群。

目前关于适应性富集设计的研究大多数集中在假设检验上,即 I 类错误控制<sup>[1-3]</sup>。然而,药物疗效估计方面的研究却相对较少<sup>[4]</sup>。尽管对药物疗效估计的关注较少,但在实际应用中,准确地估计疗效仍然非常重要。传统的估计方法是极大似然估计(maximum likelihood estimation, MLE),该方法应用简单的加权平均计算。然而,在富集设计的背景下,由于期中分析中存在亚组选择的问题,这种方法的估计结果往往存在偏倚和局限性<sup>[5]</sup>。在既往的文献研究中, Kimani 等人推导了基于 Rao-Blackwell 定理的条件无偏估计

量,并考察了该方法在期中决策条件下的统计性能<sup>[6]</sup>。Stallard 等人评估了极大似然方法的偏倚,并提出了偏差调整估计量<sup>[7]</sup>。Bowden 等学者比较了极大似然方法、收缩估计方法等不同方法,并通过偏倚和均方根误差评价了各个方法的表现<sup>[8]</sup>。

因此,如何准确地估计药物治疗效果是一个极具现实意义的统计学问题。需要考虑能否在不以均方误差膨胀为代价的情况下,改善极大似然估计所带来的偏倚。例如,第二阶段估计方法和条件无偏估计方法的偏倚较小,但第二阶段方法仅利用了部分数据;Lindley 估计方法是基于贝叶斯模型来进行估计药物疗效的,Bootstrap 方法通过赋予不同权重来进行计算,它们在不同分布类型的终点上均能减小偏倚等<sup>[9]</sup>。

本研究的目的是在给定的适应性富集设计框架下,评价不同疗效估计方法的统计学表现。通过模拟多种情形,比较不同方法的估计效果,为实际临床试验中的药物疗效点估计提供理论支持。

### 理论方法

为简化研究情形,本研究以两阶段双臂试验为例,样本量按 1:1 比例在试验组和对照组间随机分配。假设各组的观测值均服从正态分布且具有

\* 基金项目:国家自然科学基金(82273738);江苏省高校“青蓝工程”

# 共同第一作者

△通信作者:于浩, E-mail: haoyu@njmu.edu.cn; 柏建岭, E-mail: baijianling@njmu.edu.cn

相同的方差  $\sigma^2$ 。

本研究定义的期中决策规则为:若第一阶段入组的全人群中,观测到的阳性亚组疗效均值  $x_+$  与阴性亚

组疗效均值  $x_-$  之差大于临界阈值  $c$ , 则仅纳入阳性亚组人群进入下一阶段; 否则, 全人群进入下一阶段试验 (图 1)。



图 1 富集设计流程图

基于上述设定, 假设阳性亚组在全人群中的流行率为  $\pi$ 。记亚组为  $g, g \in \{+, -\}$ , 其中 + 表示阳性亚组, - 表示阴性亚组。记第一阶段和第二阶段亚组  $g$  内各处理组的样本量分别为  $n_g^{(1)}$  和  $n_g^{(2)}$ , 药物疗效 (即试验组与对照组均值之差) 的均值分别为  $X_g^{(1)}$  和  $X_g^{(2)}$ , 故有  $X_g^{(i)} \sim N(\Delta_g, 2\sigma^2/n_g^{(i)})$ ,  $i=1, 2$ , 其中  $\Delta_g$  为真实疗效且未知。基于上述规则与标识, 我们将分别介绍 5 种不同的效应估计方法。

1. 第二阶段估计方法

该方法仅使用第二阶段的数据来估计药物疗效, 具体如下:

$$\hat{\Delta}_g^{Stage2} = X_g^{(2)}$$

在第二阶段试验时并没有进行组别筛选, 因此第二阶段的效应均值可以看作是亚组疗效的无偏估计。

2. 极大似然估计方法

在期中分析时, 信息时间  $t = n_g^{(1)} / (n_g^{(1)} + n_g^{(2)})$ 。那么, 极大似然估计量如下:

$$\hat{\Delta}_g^{MLE} = tX_g^{(1)} + (1-t)X_g^{(2)}$$

虽然极大似然方法是最常用的估计方法, 但该方法存在偏倚。

3. Lindley 估计方法

Lindley 估计方法基于贝叶斯层次模型 (Bayesian hierarchical model) 对观测疗效进行建模。已知在第一阶段时, 亚组疗效均值  $X_g^{(1)} \sim N(\Delta_g, 2\sigma^2/n_g^{(1)})$ , 记  $\tilde{X}_g^{(1)} = \sqrt{n_g^{(1)}/2} \Delta_g$ , 则  $\tilde{X}_g^{(1)} \sim N(\sqrt{n_g^{(1)}/2} \Delta_g, \sigma^2)$ 。进一步假设  $\tilde{\Delta}_g = \sqrt{n_g^{(1)}/2} \Delta_g \sim N(\mu, \theta^2)$ , 可以得到:  $\tilde{X}_g^{(1)} | \mu \sim N(\mu, \theta^2 + \sigma^2)$ , 其中  $\mu$  和  $\theta^2$  为超参数<sup>[10-11]</sup>。

可通过极大似然方法来估计  $\mu$  和  $\theta$ , 即:

$$\hat{\mu} = \frac{\tilde{X}_+^{(1)} + \tilde{X}_-^{(1)}}{2}$$

$$\hat{\theta}^2 = (\tilde{X}_+^{(1)} - \hat{\mu})^2 + (\tilde{X}_-^{(1)} - \hat{\mu})^2 - \sigma^2$$

由于  $\hat{\theta}^2$  应非负数, 所以定义  $\hat{\theta}^2 = \max(0, (\tilde{X}_+^{(1)} - \hat{\mu})^2 + (\tilde{X}_-^{(1)} - \hat{\mu})^2 - \sigma^2)$ 。

基于此, 可得到后验分布  $f(\tilde{\Delta}_g | \tilde{X}_g^{(1)})$ , 其期望为:

$$E(\tilde{\Delta}_g | \tilde{X}_g^{(1)}) = \frac{\sigma^2}{\hat{\theta}^2 + \sigma^2} \hat{\mu} + \frac{\hat{\theta}^2}{\hat{\theta}^2 + \sigma^2} \tilde{X}_g^{(1)}$$

因此, 第一阶段亚组的疗效估计值为:  $\hat{\Delta}_g =$

$$\sqrt{2/n_g^{(1)}} \tilde{\Delta}_g = \sqrt{2/n_g^{(1)}} \left( \frac{\sigma^2}{\hat{\theta}^2 + \sigma^2} \hat{\mu} + \frac{\hat{\theta}^2}{\hat{\theta}^2 + \sigma^2} \tilde{X}_g^{(1)} \right), \text{ 故:}$$

$$\hat{\Delta}_g^{Lindley} = t \sqrt{2/n_g^{(1)}} (\hat{B} \hat{\mu} + (1-\hat{B}) \tilde{X}_g^{(1)}) + (1-t) X_g^{(2)}$$

记  $\hat{B} = 1/(\hat{\theta}^2 + \sigma^2)$ , 则  $\hat{B} = [\max(\sigma^2, (\tilde{X}_+^{(1)} - \hat{\mu})^2 + (\tilde{X}_-^{(1)} - \hat{\mu})^2)]^{-1}$ 。从估计量公式可以发现, 该方法将第一阶段各亚组疗效估计量收缩到共同的均值上, 在第二阶段仍然采用极大似然方法合并估计量, 因此该方法也称为收缩估计。

4. Bootstrap 参数方法

Bootstrap 方法是在第一阶段对阳性亚组和阴性亚组赋予不同的权重系数, 从而估计药物疗效<sup>[9]</sup>。在这种情况下, 可以给最有效的治疗组分配更大的权重, 然后再进行线性组合估计药物疗效。根据前面的假设,  $\Delta_g \sim N(X_g^{(1)}, 2\sigma^2/n_g^{(1)})$ 。本研究主要关注生物标志物阳性亚组和阴性亚组, 因此  $\Delta_+ - \Delta_- \sim N(X_+^{(1)} - X_-^{(1)}, 2\sigma^2/n_+^{(1)} + 2\sigma^2/n_-^{(1)})$ 。首先, 估计两个亚组中每个亚组具有较大预期治疗效果的概率, 定义概率为:  $\hat{p}_+ = \Pr(\Delta_+ > \Delta_-) = \Pr(\Delta_+ - \Delta_- > 0)$ 。那么:

$$\hat{p}_+ = \Phi \left( \frac{X_+^{(1)} - X_-^{(1)}}{\sqrt{2\sigma^2/n_+^{(1)} + 2\sigma^2/n_-^{(1)}}} \right)$$

并且:  $\hat{p}_- = 1 - \hat{p}_+$ 。根据实际数据计算的累积分布函数记为  $\hat{p}_+$ , 若两亚组疗效相等, 那么  $\hat{p}_+ = \hat{p}_- = 1/2$ , 即两个亚组的概率相等; 否则, 疗效更大的亚组分配更大的概率。然后, 通过以下步骤进行采样:

①从  $N(0, 2\sigma^2/n_+^{(1)} + 2\sigma^2/n_-^{(1)})$  分布中采样一个值  $\Delta^*$ , 代表阳性亚组和阴性亚组疗效差异的平均值;

②在标准正态分布下, 通过  $\Delta^*$  计算概率密度值:

$$f^* = \varphi \left( \frac{\Delta^*}{\sqrt{2\sigma^2/n_+^{(1)} + 2\sigma^2/n_-^{(1)}}} \right);$$

③步骤①~②重复 10000 次。

通过以上步骤可以得到  $f^*$  的样本分布, 然后将实际数据获得的  $f = \varphi \left( \frac{X_+^{(1)} - X_-^{(1)}}{\sqrt{2\sigma^2/n_+^{(1)} + 2\sigma^2/n_-^{(1)}}} \right)$  与抽样结果

进行比较。定义:  $\eta = \Pr(f^* > f)$ , 当  $\Delta_+ - \Delta_-$  存在较大疗效时,  $f$  就会较小,  $\eta$  会趋近于 1。基于此, 疗效估计量可以写为以下表达式:

$$\hat{\Delta}_{G^{(1)}}^{Bootstrap} = t \sum_{g \in \{+, -\}} (\eta \hat{p}_g + (1 - \eta) \pi_g) X_g^{(1)} + (1 - t) X_{G^{(1)}}^{(2)}$$

$$\hat{\Delta}_{G^{(2)}}^{Bootstrap} = t \sum_{g \in \{+, -\}} (\eta (1 - \hat{p}_g) + (1 - \eta) \pi_g) X_g^{(1)} + (1 - t) X_{G^{(1)}}^{(2)}$$

其中,  $G^{(i)}$  表示亚组疗效的顺序统计量 ( $i = 1, 2$ ),  $i = \{1, 2\} \rightarrow \{+, -\}$ 。若阳性亚组疗效均值更大, 那么  $G^{(1)}$  表示阳性亚组,  $G^{(2)}$  表示阴性亚组; 否则, 顺序相反。从上式可以看出, 如果阳性亚组和阴性亚组疗效相同, 则第一阶段的估计量趋近于  $t \sum_{g \in \{+, -\}} \pi_g X_g^{(1)}$ , 符合设定的情形。

### 5. 条件无偏估计方法

该方法基于 Rao-Blackwell 定理, 旨在推导出一致最小方差无偏估计量<sup>[12]</sup>。Rao-Blackwell 定理从一个无偏估计为起点, 寻找其在给定条件下具有最小方差的无偏估计量<sup>[13]</sup>。不论真实参数的取值如何, 最小方差无偏估计都比其他无偏估计具有更小或相等的方差<sup>[14]</sup>。当阳性亚组和阴性亚组均进入第二阶段时, 第二阶段的样本疗效可以视为无偏估计。若仅有阳性亚组进入第二阶段, 即:  $d(X_g^{(1)}) = \{+\}$ , 定义  $Q$  为:  $\{X_+^{(1)} > X_-^{(1)} + c\}$ ,  $b = X_-^{(1)} + c$ ; 若全人群进入第二阶段, 即:  $d(X_g^{(1)}) = \{+, -\}$ , 则定义  $Q$  为:  $\{X_+^{(1)} < X_-^{(1)} + c\}$ ,  $b' = X_+^{(1)} - c$ 。因此, 条件无偏估计量如下, 对阳性亚组的疗效估计:

$$\hat{\Delta}_+^{UMVCUE} = \begin{cases} \frac{\sigma_{2,+}^2 x_+ + \sigma_{1,+}^2 y_+}{\sigma_{1,+}^2 + \sigma_{2,+}^2} - \frac{\sigma_{2,+}^2}{\sqrt{\sigma_{1,+}^2 + \sigma_{2,+}^2}} \frac{\varphi(W_+)}{\Phi(W_+)} & d(X_g^{(1)}) = \{+\} \\ \frac{\sigma_{2,+}^2 x_+ + \sigma_{1,+}^2 y_+}{\sigma_{1,+}^2 + \sigma_{2,+}^2} + \frac{\sigma_{2,+}^2}{\sqrt{\sigma_{1,+}^2 + \sigma_{2,+}^2}} \frac{\varphi(W_+)}{1 - \Phi(W_+)} & d(X_g^{(1)}) = \{+, -\} \end{cases}$$

其中,

$$W_+ = \frac{1}{\sigma_{1,+}^2} \left( \frac{\sigma_{2,+}^2 x_+ + \sigma_{1,+}^2 y_+}{\sqrt{\sigma_{1,+}^2 + \sigma_{2,+}^2}} - b \sqrt{\sigma_{1,+}^2 + \sigma_{2,+}^2} \right) = \frac{\sqrt{\sigma_{1,+}^2 + \sigma_{2,+}^2}}{\sigma_{1,+}^2} (Z_{MLE}^+ - b)$$

类似地, 对阴性亚组的疗效估计为:

$$\hat{\Delta}_-^{UMVCUE} = \begin{cases} X_-^{(1)} & d(X_g^{(1)}) = \{+\} \\ \frac{\sigma_{2,-}^2 x_- + \sigma_{1,-}^2 y_-}{\sigma_{1,-}^2 + \sigma_{2,-}^2} - \frac{\sigma_{2,-}^2}{\sqrt{\sigma_{1,-}^2 + \sigma_{2,-}^2}} \frac{\varphi(W_-)}{\Phi(W_-)} & d(X_g^{(1)}) = \{+, -\} \end{cases}$$

其中,  $W_- = \frac{1}{\sigma_{1,-}^2} \left( \frac{\sigma_{2,-}^2 x_- + \sigma_{1,-}^2 y_-}{\sqrt{\sigma_{1,-}^2 + \sigma_{2,-}^2}} - b' \sqrt{\sigma_{1,-}^2 + \sigma_{2,-}^2} \right) = \frac{\sqrt{\sigma_{1,-}^2 + \sigma_{2,-}^2}}{\sigma_{1,-}^2} (Z_{MLE}^- - b')$ ,  $b' = X_+^{(1)} - c$

$\sigma_{1,g}^2 = 2\sigma^2/n_g^{(1)}$ ,  $\sigma_{2,g}^2 = 2\sigma^2/n_g^{(2)}$  分别表示第一阶段和第二阶段亚组疗效均值的方差。

## 模拟试验

### 1. 模拟参数设置及评价指标

使用 R 4.3.1 软件进行数据模拟和分析, 以评估不同估计方法的性能特征。本研究模拟了阴性亚组疗效均值  $\Delta_-$  设定为 0, 阳性亚组疗效均值  $\Delta_+$  为 0、0.25 和 0.5, 其中阳性亚组具有更高的治疗效果。假设方差  $\sigma^2 = 1$ , 疗效界值设定为  $c = 0.1$ 。在模拟试验中, 若亚组疗效差大于界值, 即阳性亚组继续试验, 仅估计阳性亚组的疗效; 否则, 同时估计阳性和阴性亚组的药物疗效。阳性亚组和阴性亚组总样本量为 600, 生物标志物阳性的流行率  $\pi$  分别设定为 0.25、0.5 或 0.75, 信息时间  $t$  取值为 0.1、0.2、 $\dots$ 、0.9。模拟试验重复进行 10000 次。

评价各个方法的估计效果主要使用偏倚 (bias) 和均方根误差 (root mean square error, RMSE)<sup>[15]</sup>:

$$bias = \frac{\sum_{i=1}^B (\hat{\theta}_i - \theta_i)}{B}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^B (\hat{\theta}_i - \theta_i)^2}{B}}$$

其中,  $B$  表示模拟次数,  $\hat{\theta}_i$  是第  $i$  次模拟的估计值,  $\theta_i$  是真实值。

### 2. 模拟结果

流行率和亚组疗效对阳性和阴性亚组进入下一阶段的可能性都有影响。例如, 在信息时间为 0.5 的情形下, 通过模拟试验得到阳性亚组和全人群进入第二阶段试验的概率 (表 1), 可以观察到不同的流行率对第二阶段选择的亚组有一定的影响, 尽管影响并不显著。在不同真实阳性亚组疗效的情形下, 虽然大多数情况下会正确选择继续第二阶段的亚组, 但也存在一定的概率选择不符合真实情况的亚组进入第二阶段。例如, 在流行率为 0.75 且阳性亚组疗效为 0 的情况下, 在临界阈值为 0.1 的条件下, 按照真实情况应该全人群进入第二阶段试验, 但实际上全人群进入第二阶段的概率仅为 0.7078。流行率为 0.25 和 0.75 的情况下, 亚组进入第二阶段试验的概率几乎相似, 可能是由于两个亚组样本量比例为 1:3 和 3:1 造成的影响。

#### (1) 估计偏倚

根据模拟试验的设定, 可以观察到以下情况: 在  $\Delta_+ = 0$  的情形下 (图 2, 第 1 列), 两个亚组的疗效差小于临界阈值 0.1, 因此需要对阳性和阴性亚组的效应进行估计; 然而, 在  $\Delta_+ = 0.25$  和 0.5 的情况下 (图 2, 第 2、3 列), 亚组的疗效差异大于临界值, 因此只需估计阳性亚组的治疗效果。总体来看, 相较于其他估计方法, 极大似然估计和 Lindley 估计方法存在明显的估计偏倚; 在不同的情形下, 第二阶段估计方法和条件

表 1 全人群/仅阳性亚组进入第二阶段的概率

	$\pi=0.25$		$\pi=0.5$		$\pi=0.75$	
	全人群	仅阳性亚组	全人群	仅阳性亚组	全人群	仅阳性亚组
$\Delta_+ = 0.00$	0.7072	0.2928	0.7360	0.2640	0.7078	0.2922
$\Delta_+ = 0.25$	0.2092	0.7908	0.1745	0.8255	0.2081	0.7919
$\Delta_+ = 0.50$	0.0174	0.9826	0.0085	0.9915	0.0175	0.9825

无偏估计方法的估计偏倚相对较稳定,几乎为无偏估计;而 Bootstrap 方法的偏倚较小,但受亚组疗效变化影响较大。模拟结果显示,在  $\Delta_+ = 0$  的情况下,分别考虑流行率为 0.25、0.5、0.75 的三种场景(图 2,第 1 列),极大似然方法、Lindley 方法和 Bootstrap 参数方法的估计偏倚明显不同,这可能是因为这些方法估计全人群的治疗效果,而第二阶段估计方法和条件无偏方法的结果几乎是无偏的。当仅估计阳性亚组的疗效

时,即在  $\Delta_+ = 0.25$  和 0.5 的场景中(图 2,第 2、3 列),极大似然方法显示明显的估计偏倚,会高估药物疗效;Lindley 方法也存在较大的偏倚。相比之下,Bootstrap 参数方法的估计偏倚较小。总体来看,在  $\Delta_+ = 0.25$  和 0.5 的情况下(图 2,第 2、3 列),随着流行率的增加,极大似然法、Lindley 估计法和 Bootstrap 参数法这三种方法的估计偏倚逐渐减小,可能是由于流行率增加导致阳性亚组样本量增加的影响。

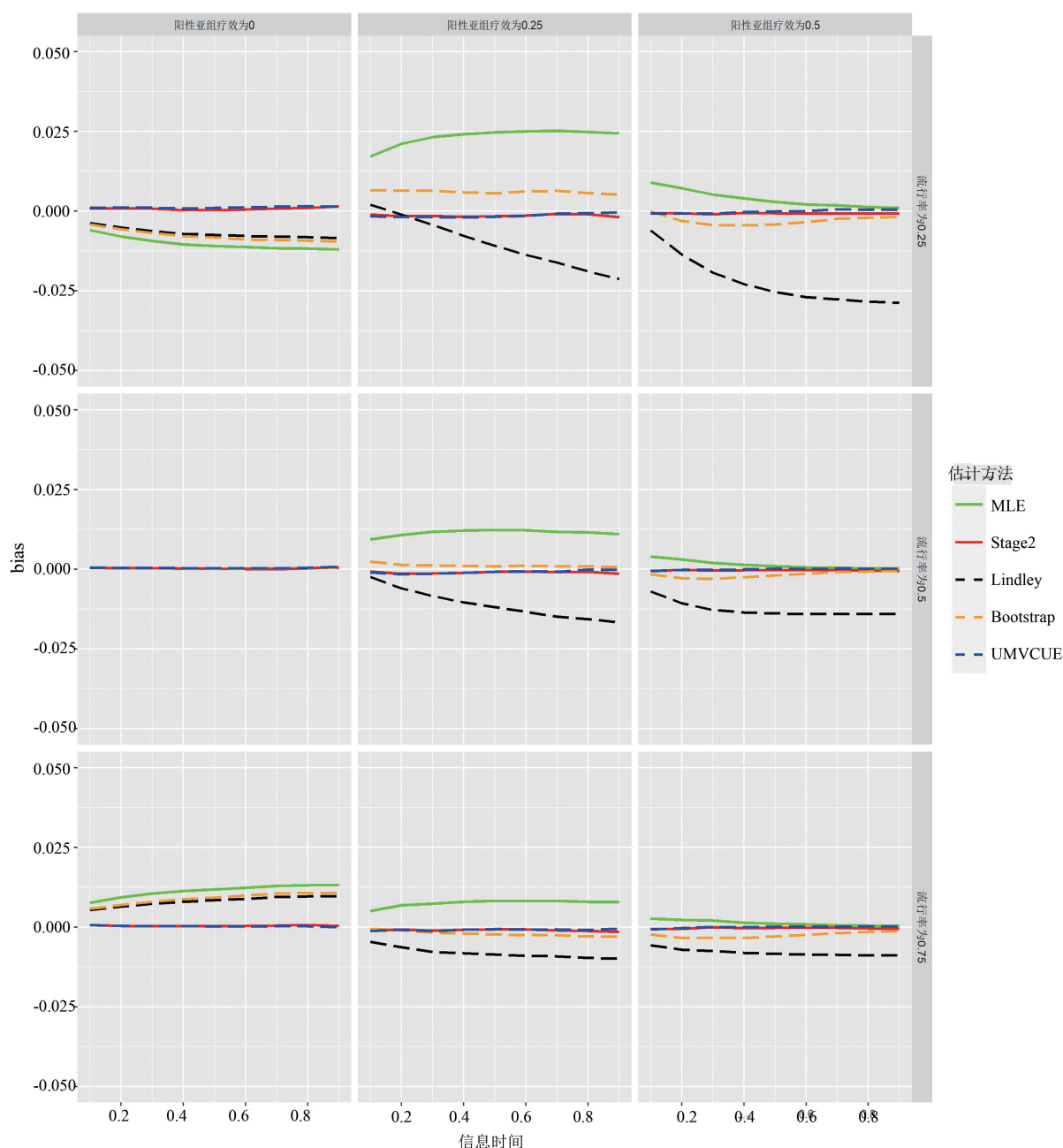


图 2 不同情形下,五种估计方法的偏倚( $\pi=0,0.25,0.5$ )

(2) 均方误差

图 3 展示了不同情况下五种估计方法的均方根误差的模拟结果。从图 3 可以观察到,在所有情形下,第二阶段估计方法的 RMSE 都较大,并且随着信息时间的增加而逐渐增大。这是因为第二阶段估计方法仅依赖于第二阶段的数据信息,并且随着期中分析的信息时间的增加,第二阶段的样本量

减少,因此容易产生较大的 RMSE。在所有情况下,其他估计方法显示出类似的 RMSE 结果。特别是在阳性亚组疗效为 0.25 和 0.5 的情况下(图 3,第 2、3 列),随着流行率的增加,所有估计方法的均方根误差都呈现下降趋势。这可能是因为随着流行率增加,阳性亚组的样本量增加,对阳性亚组疗效的估计变得更为准确和可靠。

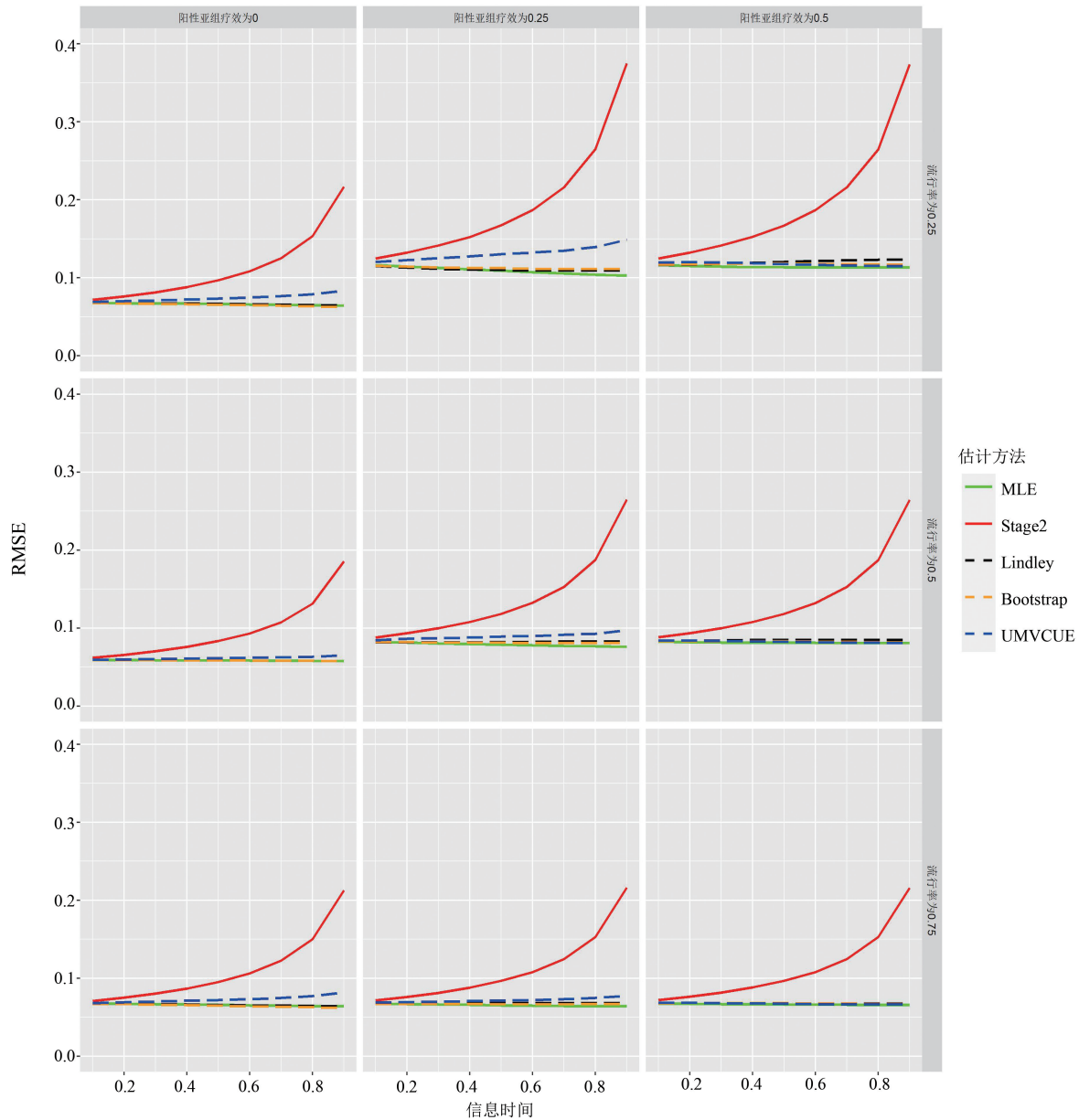


图 3 不同情形下,五种估计方法的均方根误差( $\pi=0,0.25,0.5$ )

实例分析

本研究使用一项治疗丙型肝炎的临床试验案例来展示上述五种方法的实际应用。MILLY 试验是一项随机、双盲、安慰剂对照的 II 期研究,研究了与白细胞介素-13 结合的人源化单克隆抗体 lebrikizumab 在控制不佳的哮喘中的有效性和安全性<sup>[16]</sup>。MILLY 试验为单阶段研究,为了说明目的,我们将其扩展到本文

研究背景下,使用血清骨膜蛋白作为第二阶段继续选择患者人群的生物标志物,根据实例参数数据产生模拟数据结果,即阳性亚组、阴性亚组疗效真值为 0.43 和 0.08,主要疗效指标是支气管舒张剂前 1 秒用力呼气容积从基线到 12 周的相对变化。那么,疗效均值如下(表 2)。根据以往文献<sup>[17]</sup>,假设两种情况下的流行度为 0.5,标准差为 0.19。此外,假设在治疗组和安慰剂组中样本量均为 100,并

在信息时间为 0.5 时进行期中分析,期中分析的临界阈值为 0.3。

表 2 不同阶段下各亚组的疗效均值

亚组	阳性亚组	阴性亚组
期中分析	0.62	0.17
最终分析	0.43	—

由于临界阈值为 0.3,因此,仅阳性亚组纳入第二阶段试验。通过以上的参数设定,就可以计算各个方法的点估计值(如表 3)。从表中可以发现:对于阳性亚组的疗效估计,第二阶段估计方法和条件无偏估计方法的结果比较接近。

表 3 不同方法下各亚组的点估计值

方法	阳性亚组	阴性亚组
极大似然估计	0.53	0.17
第二阶段估计	0.43	0.17
Lindley 估计	0.52	0.22
Bootstrap 估计	0.50	0.13
条件无偏估计	0.45	0.17

## 讨 论

本研究通过模拟试验发现,阳性亚组的流行率对于极大似然方法、Lindley 方法和 Bootstrap 参数方法的疗效估计结果产生影响。随着流行率的增加,这三种方法的估计偏倚明显减小。虽然第二阶段估计方法和条件无偏估计方法几乎都表现为无偏估计,但第二阶段估计方法是以均方根误差膨胀为代价的。模拟结果还显示,改进的方法在不同场景下都能或多或少地改善极大似然方法的估计偏倚。根据食品药品监督管理局相关指南<sup>[18]</sup>,适应性设计的期中决策规则可能导致观察到的治疗效果有随机偏高的现象,从而影响最终的疗效估计结果<sup>[19]</sup>。因此,在实际临床试验应用中,准确估计药物的疗效显得尤为重要。目前常用的疗效估计方法为极大似然估计,但由于其忽略了适应性调整的影响,因此在适应性富集设计中的应用往往会引起估计偏倚的问题<sup>[20]</sup>。

在适应性富集设计的背景下,本文主要关注阳性亚组和阴性亚组,通过设定不同的亚组疗效值和流行率,利用偏倚和均方根误差这两个评价指标来比较不同方法的估计效果。研究发现,在阳性亚组疗效为 0.25、0.5 时,随着流行率的增加,Bootstrap 方法和 Lindley 方法的估计偏倚逐渐减小,条件无偏估计方法的均方根误差也呈现下降趋势;而当两亚组疗效差异增大时,极大似然方法的估计偏倚逐渐减小。综上所述,在所有情形中,条件无偏估计方法的偏倚都几乎是无偏的,并且与第二阶段估计方法相比,均方根误差也较小,所以推荐使用条件无偏估计方法。总体而言,第

二阶段估计方法虽然估计偏倚较小,但其均方根误差较大,因此在实际应用中,选择合适的效应估计方法应根据具体需求来决定。

本文考虑了一个两阶段的适应性富集设计,其结果假设为正态分布,并且制定了预先指定的决策规则,根据观察到的患者治疗效果差异来选择进入下一阶段的亚组。期中决策规则基于比较亚组疗效差与预设的临界阈值大小,因此不涉及一类错误的问题。临界阈值根据临床实践确定,代表研究者希望阳性亚组疗效优于阴性亚组疗效的预期效果,因此在下一阶段仅研究阳性亚组人群。此外,本文讨论了五种效应估计方法,其他适应性设计中的估计方法也可推广至适应性富集设计中。本研究还涉及效应点估计方法及其构建可信区间的问题,这是值得进一步研究的方向。

## 参 考 文 献

- [1] Thall PR. Adaptive Enrichment Designs in Clinical Trials[J]. Annu Rev Stat Appl, 2021,8(1):393-411.
- [2] Wang SJ, Hung HM. Adaptive enrichment with subpopulation selection at interim: Methodologies, applications and design considerations[J]. Contemporary Clinical Trials, 2013,36(2):673-681.
- [3] Lin R, Yang Z, Yuan Y, et al. Sample size re-estimation in adaptive enrichment design [J]. Contemporary Clinical Trials, 2021, 100: 106216.
- [4] Stallard N. Adaptive Enrichment Designs with a Continuous Biomarker[J]. Biometrics, 2023(1):9-19.
- [5] Di Stefano F, Pannaux M, Correges A, et al. A comparison of estimation methods adjusting for selection bias in adaptive enrichment designs with time-to-event endpoints[J]. Statistics in Medicine, 2022, 41(10):1767-1779.
- [6] Kimani PK, Todd S, Stallard N. Estimation after subpopulation selection in adaptive seamless trials[J]. Statistics in Medicine, 2015,34(18):2581-2601.
- [7] Stallard N, Todd S. Point estimates and confidence regions for sequential trials involving selection[J]. Journal of Statistical Planning & Inference, 2005,135(2):402-419.
- [8] Bowden J, Brannath W, Glimm E. Empirical Bayes estimation of the selected treatment mean for two-stage drop-the-loser trials: a meta-analytic approach[J]. Statistics in Medicine, 2014,33(3):388-400.
- [9] Pickard MD, Chang M. A Flexible Method Using a Parametric Bootstrap for Reducing Bias in Adaptive Designs With Treatment Selection[J]. Statistics in Biopharmaceutical Research, 2014,6(2):163-174.
- [10] Carreras M, Brannath W. Shrinkage estimation in two-stage adaptive designs with midtrial treatment selection[J]. Statistics in Medicine, 2013,32(10):1677-1690.
- [11] Hwang JT. Empirical Bayes Estimation for the Means of the Selected Populations[J]. Sankhyū: The Indian Journal of Statistics, Series A (1961-2002), 1993(2):285-304.
- [12] Tappin L. Unbiased estimation of the parameter of a selected binomial population[J]. Communications in Statistics: Theory and Methods, 1992(4):1067-1083.

(下转第 49 页)