◇ 综述与讲座 ◇

http://www.cjcpt.com
2026 Aug; 31(8): 1101-1111

协变量适应性随机化的临床试验统计分析方法介绍

杨雨轩[1]，朱思宇[1]，尹明港[1]，潘建红[2]，段重阳[1]

[1]南方医科大学，多器官损伤防治全国重点实验室，广州 510515，广东；
[2]国家药品监督管理局药品审评中心，北京 100022

摘要　因可以平衡试验组间可能对试验结果产生影响的预后因素，协变量适应性随机化目前被越来越多地应用于药物临床试验之中。而在进行此类试验的统计分析时，研究者通常认为需要对随机化控制的协变量进行一定调整，因此回归模型以及分层分析常被用于该类试验的统计推断。然而，尚未有研究系统讨论上述方法的合理性以及适用性。为此，本研究对既往文献进行了梳理，讨论了传统统计推断方法在协变量适应性随机化临床试验中的适用性，并对近些年提出的“稳健”统计推断方法进行了介绍与探讨。此外，为进一步方便临床研究人员的使用，本文还对所讨论的方法进行了实例代码演示。

关键词　适应性随机化；分层分析；协变量调整；稳健分析方法；软件操作

中图分类号：R195.1；R969.4
文献标志码：A
文 章 编 号：1009-2501(2026)08-1101-11
doi: 10.12092/j.issn.1009-2501.2026.08.011

在随机对照临床试验中，随机分配是确保治疗效应归因准确性与推断可靠性的核心机制，也是实现药物效应定量推断和控制偏倚的基础手段之一[1]。尽管简单随机化在理论上被认为是一种直接且有效的分配方法，但在实际应用中，可能导致重要协变量在组间分布不平衡，从而影响效应估计的准确性。为降低该风险，当前更倾向于在随机化过程中对关键协变量进行控制，主要方法包括分层区组随机化[1]和协变量适应性随机化[2]。

对于采用分层随机化的试验，目前已形成共识：在分析阶段应将分层因素纳入统计模型。这可以通过分层检验方法（如分层 Mantel-Haenszel 检验、分层 Log-Rank 检验）实现，或通过将分层因素作为协变量引入模型（如一般线性模型、Logistic 回归、Cox 回归等）来完成。对于协变量适应性随机化设计的试验，已有研究指出，若在分析中忽略协变量的影响，可能导致估计偏倚及检验效能下降。尽管已有学者针对协变量纳入分析的策略提出了多种方法[3-7]，目前仍缺乏对这些方法的系统整理与比较。

本研究旨在系统梳理现有针对协变量适应性随机化试验的统计推断方法，为后续临床试验的设计与分析提供理论支撑与实践参考。文章首先介绍常见的协变量适应性随机化方法，阐述其理论基础与实现机制；随后系统评述用于不同数据类型的统计推断方法，包括传统方法与稳健方法；最后，通过具体实例与相关代码演示各类方法的实际应用，为临床研究者提供具有可操作性的指导。

1　方法

1.1　协变量适应性随机化方法　协变量适应性随机分配是指当前受试者的分配，部分或完全取决于其自身的基线特征以及已入组受试者的基线特征和随机分配结果[1]。本文首先给出分层区组随机化方法，然后给出主要讨论的协变量适应性随机化方法，包括协变量适应性偏置硬币随机化、分层瓮设计随机化和最小化随机化[8]。

1.1.1　分层区组随机化　先按重要基线特征对受试者分层，然后在每层内再进行独立的随机分配，

2025-04-21 收稿　2025-06-16 修回
国家自然科学基金（82273727）；广州市科技计划项目基金（2023A04J2266）
杨雨轩，男，研究方向：临床试验设计与分析。
E-mail: yuxuanyang402@qq.com
潘建红，通信作者，男，副研究员，主要从事药品的技术审评工作。
E-mail: panjh@cde.org.cn
段重阳，通信作者，男，博士，副教授，硕士生导师，研究方向：临床试验方法学及临床研究。
E-mail: donyduang@126.com

即为分层随机。在分层基础上，如果各层内采用区组随机分配，则被称为分层区组随机化[1,9]。分层随机化可保证分层因素在组间的均衡性，区组随机化则可使同一时间段同一区组内的受试者在各个治疗组间的分配比例符合预设要求，分层区组随机化可以兼具二者的优点。但分层随机化会随着分层因素的增多，随机化的层数呈指数级增长，可能导致某些层受试者过少，影响统计分析。因此当分层因素较多时，多考虑采用协变量适应性随机化方法。

1.1.2　协变量适应性偏置硬币随机化　协变量适应性偏置硬币随机化根据试验中组间受试者协变量的均衡情况，调整后续受试者的治疗分配概率，以达到平衡协变量的目的。具体而言，对于一两组 1:1 设计，若使用a_1表示试验组，a_0表示对照组，则对于第$i, i \in \{1,2,\ldots,n\}$个受试者，根据其$k$维协变量$Z_i = (z_{i1}, z_{i2}, \ldots, z_{ik})$，其分配至治疗组的概率$P(A_i = a_1)$可表示为[10]：

$$P(A_i = a_1) = \begin{cases} p, & D_{i-1}(z_i) < 0, \\ 1/2, & D_{i-1}(z_i) = 0, \\ 1-p, & D_{i-1}(z_i) > 0 \end{cases} \quad (1)$$

其中 p 为事先指定的分配概率，满足$1/2 < p < 1$，而$D_{i-1}(z_i)$为在协变量 Z 的联合水平上$(z_{i1}, z_{i2}, \ldots, z_{ik})$前$i-1$位受试者被分配至试验组的人数与分配至对照组人数之差，用于表示前$i-1$位受试者在协变量Z联合水平上的均衡情况，为不失一般性，定义$D_0 = 0$。对于多治疗组的协变量适应性偏置硬币随机化方法可参考 Kuznetsova 等的研究[11]。

1.1.3　分层瓮设计随机化　分层瓮设计随机化属于偏置硬币随机化，只是其中的概率 p 会进行适应性变化。具体而言，对于上述 1:1 设计下的第$i, i \in \{1,2,\ldots,n\}$个受试者，其分配至治疗组的概率$P(A_i = a_1)$为：

$$P(A_i = a_1) = \begin{cases} p_i, & D_{i-1}(z_i) < 0, \\ 1/2, & D_{i-1}(z_i) = 0, \\ 1-p_i, & D_{i-1}(z_i) > 0 \end{cases} \quad (2)$$

其中p_i根据下式进行计算：

$$p_i = 1/2 + \omega|D_{i-1} - 1|/[2s + \omega(i-1)]$$

式中ω和 s 均为事先指定的非 0 实数。在这种情况下，瓮设计在随机化的开始阶段便会强制减少治疗组间协变量的不平衡，并随着试验样本量的增加，趋近简单随机[12-13]。

1.1.4　最小化随机化　不同于上述协变量适应性偏置硬币随机化和分层瓮设计随机化使用协变量Z的联合水平对分配概率进行调整，最小化随机化使用的是协变量Z的边际水平，因此该方法又被称为边际方法[14-15]。具体而言，对于上述 1:1 设计下的第$i, i \in \{1,2,\ldots,n\}$个受试者，若令$G_i^{(1)}$表示其分配到试验组时，两治疗组在协变量$Z$边际水平上的不平衡度，令$G_i^{(0)}$表示其分配到对照组时两治疗组的不平衡度，则根据事先指定的分配规则，受试者i分配到a_j组的概率$P(A_i = a_1)$表示为：

$$P(A_i = a_1) = \begin{cases} p, & G_i^{(1)} < G_i^{(0)}, \\ 1/2, & G_i^{(1)} = G_i^{(0)}, \\ 1-p, & G_i^{(1)} > G_i^{(0)} \end{cases} \quad (3)$$

其中 p 为事先指定的分配概率，满足$1/2 < p < 1$，而$G_i^{(j)}, j \in \{0,1\}$可根据下式计算：

$$G_i^{(j)} = \sum_{k'=1}^{k} D_{i,k'}(A_i = a_j, Z = z_{ik'}) \quad (4)$$

其中$D_{i,k}(A_i = a_j, Z = z_{ik})$定义为第$i$位受试者被分配到治疗组 j 时，治疗组在协变量Z边际水平$(Z = z_{ik})$上的k维协变量的不平衡度之和[16]。显然，当协变量Z为一维，即$k = 1$时，最小化随机化等价于协变量适应性偏置硬币随机化；而当协变量 Z 为多维时，最小化随机化使用协变量Z的边际水平与分层偏置硬币随机化使用协变量Z的联合水平会产生很大差别。

1.2　传统统计推断方法　本文中传统统计推断方法指在简单随机化和分层随机化下开发并广泛应用的统计推断方法。为评估方法的适用性，本研究参考 Shao 和 Ye 等的研究[3-8]，对统计推断方法的“有效性”做出如下定义：对于给定的原假设H_0和双侧检验水准$\alpha > 0$，当检验统计量 T 落在区间$[z_{\alpha/2}, z_{1-\alpha/2}]$之外时拒绝原假设$H_0$。这里的$z_r$表示一个已知分布的 r 分位数，通常这个分布近似为正态分布。则若在H_0成立时，存在：

$$\lim_{n\to\infty} P(T \notin [z_{\alpha/2}, z_{1-\alpha/2}]) \le \alpha \quad (5)$$

则称检验统计量T是渐进有效的。

而若存在常数$\alpha_0 < \alpha$，使得：

$$\lim_{n\to\infty} P(T \notin [z_{\alpha/2}, z_{1-\alpha/2}]) \le \alpha_0 \quad (6)$$

则称检验统计量 T 是渐进保守的。

在使用协变量适应性随机化的临床试验中，

我们首先关心的问题是，使用基于简单随机化提出的检验方法，即完全不考虑协变量的方法，是否是有效的。其次关心的问题是，使用基于分层随机化提出的检验方法，即纳入协变量的一般线性模型、Logistic 模型、Cox 模型，或是分层分析的方法如分层 Log-Rank 检验等，是否是有效的。我们将在下文中按不同类型的结局变量讨论。

1.2.1　*连续型结局变量情形下的传统统计推断方法*　当我们研究的主要结局变量为近似正态分布的连续型变量时，不考虑协变量的传统方法是两样本*t*检验，2010 年 Shao 等证明在采用协变量适应性偏置硬币随机化的情形下，采用两独立样本 *t* 检验是保守的[3]。同样地，传统检验在分层区组随机化和分层瓮设计随机化中也是保守的[8]。而在最小化随机化情形下，Ma 等[17]在给定结局变量Y_i和用于随机化的离散型协变量Z_i的关系是完全线性且所有Z_i相互独立的条件下，证明了两独立样本 *t* 检验是保守的，但这个条件是非常局限的且在实际情形下难以满足。在更广泛适用的条件下，最小化随机化中传统检验方法是否有效这一问题还没有理论研究将其完全解决。

虽然，传统检验方法在大部分的协变量适应性随机化方法中的一类错误已被证明可以被合理控制，但是 2018 年 Bugni 等[6]通过对两样本 *t* 检验的理论推导和模拟研究认为一类错误的过保守，很可能导致检验效能的损失，而 Ma 等[18]也认为由于协变量用于随机化产生了更平衡的分配，从模型中排除这些协变量会导致检验统计量的抽样分布不真实，从而导致无效的统计推断。因此，需要在检验方法中考虑协变量。

目前对于连续型结局变量，常用的考虑协变量的方法为协方差分析（analysis of covariance, ANCOVA）或一般线性模型。早期统计模拟研究显示最小化随机化下协方差分析能够控制一类错误[17]。而近期的理论研究表明[6]，协方差分析能够控制一类错误并达到最大检验效能的前提是将适应性随机化中的所有分层因素（分层变量均为离散变量）全部纳入模型。并且，当治疗分组为两组时，协方差分析无论模型假设正确与否，协方差分析都比方差分析具有更高的检验效能。但当治疗分组为三组及以上时，协方差分析虽然在纳入所有分层因素时依旧有效，但其检验效能大大降低，且有可能低于不考虑任意协变量的方差分析。而对于一般线性模型，Ye[4]指出，只有在模型假定正确时，传统的 Wald 检验才是有效的。

Shao 等在协变量适应性随机化下提出一个服从正态分布或 *t* 分布的检验统计量*T*能有效控制一类错误需要满足的两个条件[3]：

① 用于协变量适应性随机化的协变量Z_i，需要作为构建 *T* 统计量的所有协变量X_i的一个函数被纳入。

② 统计量 *T* 对任意给定的治疗分组A_i的集合都是有效的。

满足条件②意味着在构建统计量 *T* 时建立了一个正确的统计模型，然而构建正确设定的模型是比较困难的，我们更可能碰到的是传统统计推断方法在协变量适应性随机化下不满足有效性的情况[8]。

1.2.2　*二分类结局变量情形下的传统统计推断方法*　Shao 等[19]在二分类结局变量下，对如下统计量T_S在协变量适应性随机化下的有效性进行了研究：

$$T_S = \frac{\widehat{p_1} - \widehat{p_0}}{\left\{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_0}(1-\widehat{p_0})}{n_0}\right\}^{\frac{1}{2}}} \tag{7}$$

其中n_1和n_0表示试验组和对照组的样本量，总样本量$N = n_1 + n_0$，$\widehat{p_1} = \sum_{i=1}^{N} A_i Y_{i1}/n_1$，$\widehat{p_0} = \sum_{i=1}^{N} A_i Y_{i0}/n_0$。他们的理论结果表明，$T_S$在协变量适应性偏置硬币随机化方法下是保守的。Ma 等[18]和 Shao 等[19]将前文所提到的两个条件[3]推广到二分类结局中，他们对 Logistic 模型下传统 Wald 检验的统计性质进行了理论研究，发现在协变量适应性偏置硬币随机化下基于 Logistic 模型的 Wald 检验在纳入所有协变量时有效，而在忽略部分协变量时保守。对于上述方法在其它类型的协变量适应性随机化下的适用性问题，目前尚缺乏理论研究。

1.2.3　*生存分析情形下的传统统计推断方法*　对于生存结局数据，若不考虑相关协变量的调整，常用的方法有 Log-Rank 检验和基于 Cox 模型的 Score 检验。若考虑协变量的调整，则常用的方法为分层 Log-Rank 检验和基于 Cox 模型调整协变量的 Score 检验。

传统的 Score 检验是基于 Cox 比例风险模型对于H_0的偏似然 Score 检验，具体而言，考虑如下的 Cox 比例风险模型：

$$\lambda(t|A_i,\ X_i) = \lambda_0(t)\exp\left(\theta A_i + \beta^T X_i\right) \quad (8)$$

其中θ和β是未知参数，$A_i = a_j$表示受试者被分配到治疗组（$A_i = a_1$）或对照组（$A_i = a_0$），X_i表示受试者的协变量水平，$\lambda_0(t)$表示基线风险函数。则基于 Cox 模型的 Score 检验为：

$$T_S = n^{-\frac{1}{2}} \frac{U_\theta\left(0,\widehat{\beta_0}\right)}{\widehat{A}\left(0,\widehat{\beta_0}\right)^{\frac{1}{2}}} \quad (9)$$

其　中　$U_\theta\left(0,\widehat{\beta_0}\right) = \frac{\partial \log\{L(\theta,\beta)\}}{\partial\theta}|_{\theta=0}$，$\widehat{A}\left(0,\widehat{\beta_0}\right) = -\frac{1}{n}\frac{\partial^2 \log\{L(\theta,\beta)\}}{\partial\theta^2}|_{\theta=0}$，其中$L(\theta,\beta)$为似然函数。

不考虑协变量即未分层的 Log-Rank 检验统计量可表示为：

$$T_L = n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^\tau \left\{A_i - \frac{\overline{Y_1}(t)}{\overline{Y}(t)}\right\} dN_i(t)/\hat{\sigma} \quad (10)$$

分层 Log-Rank 检验可表示为：

$$T_{SL} = n^{-\frac{1}{2}} \sum_z \sum_{i:Z_i=z} \int_0^\tau \left\{A_i - \frac{\overline{Y_{1z}}(t)}{\overline{Y_z}(t)}\right\} dN_i(t)/\hat{\sigma} \quad (11)$$

Ye 等[5]在协变量适应性随机化下研究了上述常用的传统统计检验方法的有效性。理论结果表明这些未考虑协变量调整或分层的传统统计检验方法在协变量适应性随机下是保守的。当满足如下两个条件时，被证明在本文讨论的随机化方法下均是有效的，包括最小化随机化：

① 用于随机化分组的协变量Z全部用于构建检验统计量。

② 基于模型的检验方法中的模型假设正确。

要满足这两个条件，在使用T_{SL}进行检验时需要在所有Z的联合水平上完全分层，才能保证T_{SL}是有效的。

1.3　“稳健”统计推断方法　近年来，针对上述传统统计推断方法无法实现渐近有效以及检验效能不足等问题，已有多位学者提出了相应的“稳健”统计推断方法，而这些方法均具有检验效能高、适用性广以及方差估计稳健的优点。以下按数据类型，对各个“稳健”统计推断方法进行介绍。

1.3.1　*连续型结局变量适用的“稳健”统计检验方法*　对于连续型结局下的“稳健”统计推断方法，2022 年 Ye 和 Shao 等基于传统协方差分析方法提出了异质协方差分析方法（analysis of heterogeneous COVAriance, ANHECOVA）[7]。具体而言，ANHECOVA 根据如下工作模型进行展开：

$$E(Y_i|A_i,X_i) = \theta^T A_i + \sum_{j=1}^{k} \beta_j^T (X_i - \mu_x) I(A_i = a_j) \quad (12)$$

其中，Y_i表示第 i 受试者的结局变量，A_i表示治疗分组，X_i表示包括用于随机化的协变量Z_i在内的所有协变量，$\theta,\beta_1,\ldots,\beta_j$为未知参数，而$I(A_i = a_j)$则为示性函数，用于指示受试者$i$所分配的治疗组别，$\mu_X = E(X_i)$表示基线协变量 X_i 的均值向量。由于上述模型包含了治疗分配A_i与协变量X_i的交互项，因此其可以用于刻画具有不同协变量的患者在接受相同治疗时获益不同的情况。

基于上述模型12，他们提出了广泛适用且可用于小样本临床试验的β_j的估计量$\widehat{\beta}_j$，并以此得到θ的估计量$\hat{\theta}_{ANHC}$。Ye 证明，在本文讨论的随机化方法的临床试验中，即便上述模型假设错误，ANHECOVA 仍可以获得与传统协方差分析相等或更高的检验效能。此外，当把所有用于随机化的协变量Z_i纳入工作模型时，ANHECOVA 方法检验效能的提升幅度会达到最大。然而，值得注意的是，Ye 指出，在未纳入全部Z_i时，对于采用最小化随机化的临床试验，使用 ANHECOVA 方法尚无法保证检验效能的提高。

1.3.2　*二分类结局适用的“稳健”统计检验方法*　对于二分类结局下的“稳健”统计推断方法，Wang 等[20]于 2023 年在协变量适应性偏置硬币随机化下提出了基于 Logistic 模型的“稳健”统计检验方法，但该方法并不能保证检验效能的提高。而后，Bannick 等[21]于 2025 年基于广义线性模型提出了增强逆倾向加权（augmented inverse probability weighting, AIPW）估计量，基于该统计量，研究者可对二分类结局下的治疗效应进行“稳健”的估计。具体而言，AIPW 估计量$\widehat{\theta_{AIPW}}$可使用如下表达式进行计算：

$$\widehat{\theta_{AIPW}} = \left(\overline{Y}_j - \frac{1}{n_j}\sum_{i:A_i=a_j} \hat{\mu}_j(X_i) + \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_j(X_i)\right)^T \quad (13)$$

其中$\overline{Y}_j$表示治疗组j中观察结局的样本均值，是治疗组均值参数$\theta_j = E(Y_j)$的一致估计，$\hat{\mu}_j(X_i)$表示对条件结局均值$E(Y_j|X_i)$的估计。由于治疗分配不影响基线协变量X的分布，$1/n_j\sum_{i:A_i=a_j}\hat{\mu}_j(X_i) - 1/n\sum_{i=1}^{n}\hat{\mu}_j(X_i)$具有渐进零均值，因而$\widehat{\theta_{AIPW}}$对于工作模型误设具有稳健性[21]。

为使得上述方法广泛适用，即适用于上节所提及的各个协变量适应性随机化临床试验，Bannick 通过拟合Y_j和$\hat{\mu}_j(X_i)$之间的线性回归模型对公式 13 中的$\hat{\mu}_j(X_i)$进行线性矫正得到估计量$\widehat{\theta_{LC}}$，又通过公式 14

$$\hat{\mu}_j^*(X_i) = \hat{\alpha}_j Z_i + \hat{\beta}_j \hat{\mu}(X_i) \tag{14}$$

将公式 13 中的$\hat{\mu}_j(X_i)$替换为$\hat{\mu}_j^*(X_i)$进行联合矫正得到估计量$\widehat{\theta_{JC}}$，其中$\hat{\alpha}_j, \hat{\beta}_j$有$(\hat{\alpha}_j, \hat{\beta}_j) = argmin_{\alpha_j,\beta_j} \sum_i \{Y_{i,j} - \alpha_j Z_i - \beta_j \hat{\mu}(X_i)\}^2$。

通过计算机模拟，Bannick 指出上述$\widehat{\theta_{JC}}$适用于上节所提及的所有随机化方法，且相较于$\widehat{\theta_{AIPW}}$，其方差更小。模拟结果显示，包括$\widehat{\theta_{AIPW}}$、$\widehat{\theta_{LC}}$和$\widehat{\theta_{JC}}$在内的所有估计量，其偏倚相对于各自的标准差均可忽略。此外，Shao 指出 ANHECOVA 也可用于二分类数据的分析[7]，通过我们的模拟研究发现，在纳入全部的分层因素和协变量到模型中时，$\widehat{\theta_{JC}}$与 ANHECOVA 在处理二分类数据时的统计效能几乎相同。

1.3.3　生存分析中的“稳健”统计检验方法　Ye 和 Shao 等[5]在 2020 年构建了稳健的 Log-Rank 检验T_{RL}和稳健 Score 检验T_{RS}，这两个稳健方法是对传统 Log-Rank 检验和 Score 检验进行方差调整得到的。在拟合模型错误设定的情况下，新提出的方差估计量和检验方法依旧具有渐进有效性和稳健性。尽管如此，T_{RL}和T_{RS}对于不同的随机化方法计算方式不同，无法满足广泛适用的要求，因而在 2024 年 Ye[22]提出了协变量调整的 Log-Rank 检验T_{CL}和协变量调整的分层 Log-Rank 检验T_{CSL}。考虑公式 10 中的 Log-Rank 检验可表示为：

$$T_L = n^{\frac{1}{2}} \frac{\widehat{U_L}}{\widehat{\sigma_L}} \tag{15}$$

其中$\widehat{U_L} = \frac{1}{n}\sum_{i=1}^{n} \int_0^{\tau} \left\{ A_i - \frac{\overline{Y_1}(t)}{\overline{Y}(t)} \right\} dN_i(t)$，并且满足如下的线性关系：

$$\widehat{U_L} = \frac{1}{n} \sum_{i=1}^{n} \{A_i O_{i1} - (1 - A_i) O_{i0}\} \tag{16}$$

其中

$$O_{ij} = \int_0^{\tau} \{1 - \mu(t)\}^j \{\mu(t)\}^{1-j} \{dN_{ij}(t) - Y_{ij}(t) p(t) dt\}, \quad j = 0,1 \tag{17}$$

考虑对O_{ij}和$\widehat{U_L}$之间加入协变量调整项拟合广义回归模型，得到

$$\begin{aligned}\widehat{U_{CL}} =& \frac{1}{n} \sum_{i=1}^{n} \left\{ A_i \left\{ O_{i1} - \left(X_i - \overline{X}\right)^T \hat{\beta}_1 \right\} - (1 - A_i) \left\{ O_{i0} - \left(X_i - \overline{X}\right)^T \hat{\beta}_0 \right\} \right\} \\ =& \widehat{U_L} - \frac{1}{n} \sum_{i=1}^{n} A_i \left\{ \left(X_i - \overline{X}\right)^T \hat{\beta}_1 \right\} - (1 - A_i) \left\{ \left(X_i - \overline{X}\right)^T \hat{\beta}_0 \right\}\end{aligned} \tag{18}$$

其中$\hat{\beta}_j = \left\{ \sum_{i:A_i=j} \left(X_i - \overline{X_j}\right)\left(X_i - \overline{X_j}\right)^T \right\}^{-1} \sum_{i:A_i=j} \left(X_i - \overline{X_j}\right) \hat{O}_{ij}$，因此得到协变量调整的 Log-Rank 检验

$$T_{CL} = n^{\frac{1}{2}} \frac{\widehat{U_{CL}}}{\widehat{\sigma_{CL}}} \tag{19}$$

其中$\widehat{\sigma_{CL}} = \widehat{\sigma_L} - \pi(1 - \pi)\left(\hat{\beta}_1 + \hat{\beta}_0\right)^T \hat{\Sigma}_X \left(\hat{\beta}_1 + \hat{\beta}_0\right)$

为进一步对分层因素进行调整，在公式 11 的基础上拟合广义回归模型进行协变量调整得到

$$T_{CSL} = n^{\frac{1}{2}} \frac{\widehat{U_{CSL}}}{\widehat{\sigma_{CSL}}} \tag{20}$$

其中

$$\begin{aligned}\widehat{U_{CSL}} =& \widehat{U_{SL}} - \frac{1}{n} \sum_z \sum_{i:Z_i=z} \left\{ A_i \left\{ \left(X_i - \overline{X_z}\right)^T \hat{\gamma}_1 \right\} - (1 - A_i) \left\{ \left(X_i - \overline{X_z}\right)^T \hat{\gamma}_0 \right\} \right\}\end{aligned} \tag{21}$$

$$\widehat{\sigma_{CSL}} = \widehat{\sigma_{SL}} - \pi(1 - \pi)\left(\hat{\gamma}_1 + \hat{\gamma}_0\right)^T \left\{ \sum_z \left(\frac{n_z}{n} \hat{\Sigma}_{X|z} \right) \right\} \left(\hat{\gamma}_1 + \hat{\gamma}_0\right) \tag{22}$$

Ye 的模拟研究显示方差调整后的 Log-Rank 检验T_{CL}相较于传统的分层 Log-Rank 检验T_{SL}，在一类错误被合理控制的前提下，具有更高的检验效能。同时，模拟研究表明数据在满足 Cox 模型等比例风险假设时T_{CSL}比T_{CL}方法有更高的检验效能，而在不满足该假设时则不一定。基于以上原因，我们认为当数据满足等比例风险假设时可以使用基于 Cox 比例风险模型的T_{CSL}，否则建议使用非参数方法T_{CL}。此外当数据出现稀疏层时T_{CSL}的检验效能也会大大降低，此时也建议改用T_{CL}。综合来讲，由于使用T_{CL}依赖的假设更少，我们推荐使用T_{CL}。

如表 1 所示，本文对当前常用的统计推断方法进行了归类与整理，涵盖其适用的结局变量类型以及在协变量适应性随机化设计下的可行性。总体可将推断方法划分为三类：传统方法、协变量调整方法以及“稳健”方法。不同方法在应对协

变量适应性随机化时的适用性和性能存在差异，需根据具体研究设计与数据特征进行选择。

表 1　统计推断方法总结
Tab.1　Summary of statistical inference methods

结局类型	传统方法	协变量调整方法	“稳健”方法
连续结局	方差分析	协方差分析	异质协方差分析方法
二分类结局	基于Logistic模型的Wald检验	-	增强逆倾向加权估计量、线性校正法、联合校正法
生存结局	Log-Rank检验、基于Cox模型的Score检验	分层Log-Rank检验、基于Cox模型调整协变量的Score检验	协变量调整的Log-Rank检验、协变量调整的分层Log-Rank检验、稳健Score检验

2　实例分析

本节将根据结局变量类型，展示传统统计推断方法与“稳健”统计推断方法的 R 代码及分析结果，所有分析均基于 R 包 RobinCar 实现，相关代码详见附录。

分析所用数据来自 RobinCar 包中的模拟数据集，包含三个治疗组（分别编码为 0、1、2），两个离散型分层协变量Z_1, Z_2，以及两个连续型协变量X_1, X_2。原始结局变量为连续型结局变量，在进行二分类结局分析时，我们将以 2 为临界值将结局变量转为二分类变量（大于 2 为 1，否则为 0）。

对于连续型结局，我们采用以下方法进行比较：传统的方差分析、协方差分析和新提出的 ANHECOVA 分析方法，在分析过程中我们将所有协变量纳入工作模型中（表 2）。

表 2　连续型结局实例结果
Tab.2　Example results for continuous outcomes

组别		方法		
		方差分析	协方差分析	异质协方差分析方法
治疗组				
0	均值估计	3.168	3.183	3.115
	标准误	0.232 1	0.239 3	0.179 6
1	均值估计	3.003	2.971	2.950
	标准误	0.092 9	0.084 3	0.081 5
2	均值估计	2.824	2.847	2.872
	标准误	0.094 5	0.091 6	0.090 5
组间比较				
0-1	均值差估计	0.165	0.211	0.165
	标准误	0.244	0.248	0.198
0-2	均值差估计	0.345	0.336	0.243
	标准误	0.247	0.255	0.205

表 2 显示各治疗组的效应估计值及其标准误，同时报告治疗组 0 分别与治疗组 1 和治疗组 2 之间的均值差估计及标准误。结果显示，相比于方差分析，协方差分析估计治疗组 1 和治疗组 2 的标准误略微减小，而估计治疗组 0 的标准误略有增加。ANHECOVA 在所有治疗组中均实现了较为显著的标准误降低，这与 Shao 等[7]的模拟研究的结果一致。

对于二分类结局，纳入下列统计分析方法进行比较：传统的方差分析、协方差分析、新提出的 ANHECOVA 分析方法以及 AIPW 方法、线性矫正的 AIPW 方法$\widehat{\theta_{LC}}$和联合矫正的 AIPW 方法$\widehat{\theta_{JC}}$。

同样地，所有协变量纳入工作模型中（表 3）。

表 3 显示各治疗组（Y_{bin} = 1）的概率估计及其标准误，同时报告治疗组 0 分别与治疗组 1 和治疗组 2 之间的概率差估计及标准误。$\widehat{\theta_{LC}}$和$\widehat{\theta_{AIPW}}$均实现了相较方差分析更低的标准误，且与协方差分析的标准误结果较为接近。而$\widehat{\theta_{JC}}$的标准误与 ANHECOVA 结果高度一致，且在所有方法中最小。Bannick 等[21]的模拟结果表明，包括$\widehat{\theta_{AIPW}}$、$\widehat{\theta_{LC}}$和$\widehat{\theta_{JC}}$在内的估计量，其偏倚相对于各自的标准差均可忽略。

表 3　二分类结局实例结果

Tab.3　Example results for binary outcomes

组别		方法					
		方差分析	协方差分析	异质协方差分析方法	$\widehat{\theta_{AIPW}}$	$\widehat{\theta_{LC}}$	$\widehat{\theta_{JC}}$
治疗组							
0	概率估计	0.553	0.558	0.552	0.548	0.545	0.560
	标准误	0.0311	0.0340	0.0289	0.0319	0.0323	0.0297
1	概率估计	0.763	0.755	0.756	0.767	0.767	0.771
	标准误	0.0278	0.0268	0.0268	0.0273	0.0273	0.0272
2	概率估计	0.712	0.719	0.722	0.712	0.712	0.711
	标准误	0.0307	0.0301	0.0301	0.0308	0.0308	0.0308
组间比较							
0-1	概率差估计	−0.210	−0.197	−0.204	−0.219	−0.222	−0.210
	标准误	0.0399	0.0416	0.0378	0.0400	0.0403	0.0384
0-2	概率差估计	−0.159	−0.161	−0.171	−0.163	−0.167	−0.155
	标准误	0.0430	0.0451	0.0418	0.0438	0.0441	0.0423

对于生存结局，我们选用 RobinCar 包中的 data_gen() 函数生成模拟数据，模拟参数设定如下：

真实风险函数为$\lambda_t = \lambda_0 \exp(\theta A + 1.5Z_1 - Z_{21} - 0.5Z_{22})$，删失时间$C \sim U(20,40)$，协变量$Z_1$为二分类变量且$P(Z_1 = 1) = 0.5$，$Z_2$也为离散型变量且$P(Z_2 = 1) = 0.4$，$P(Z_2 = 2) = 0.3$，$P(Z_2 = 3) = 0.3$，$Z_1$和$Z_2$为用于随机化的离散型协变量且删失时间 C 与 (A, Z_1, Z_2) 之间独立。模拟产生 200 例样本，θ设置为 0。

本文比较以下六种统计方法：传统 log-rank 检验T_L，分层 log-rank 检验T_{SL}，Score 检验T_S，“稳健”Score 检验T_{RS}，协变量调整的 log-rank 检验T_{CL}和协变量调整的分层 log-rank 检验T_{CSL}（表 4）。

生存数据实例结果如表 4 所示，T_L和T_S的 P 值在几种统计检验方法中最大，相比之下，分层 Log-Rank 检验T_{SL}的 P 值略有降低，而新方法“稳健”Score 检验T_{RS}和协变量调整的分层 Log-Rank 检验T_{CSL}的 P 值明显降低。在本实例中，协变量调整的 Log-Rank 检验T_{CL}在实例中 P 值是最小的。

表 4　生存数据实例结果

Tab.4　Example results for survival data

方法	P 值
T_L	0.400
T_{SL}	0.289
T_S	0.400
T_{RS}	0.269
T_{CL}	0.206
T_{CSL}	0.288

T_L：传统Log-Rank检验，T_{SL}：分层Log-Rank检验，T_S：Score检验，T_{RS}：“稳健”Score检验，T_{CL}：协变量调整的Log-Rank检验，T_{CSL}：协变量调整的分层Log-Rank检验。

3 小结及展望

随着协变量适应性随机化方法在临床试验中的广泛应用，如何在分析阶段对数据进行科学有效的处理，已成为方法学研究和实际应用中亟须关注的重要问题。本文围绕如何在保证推断有效性的同时提升统计检验效能的问题，对当前相关研究成果进行了系统梳理与分析。总体而言，在分层区组随机化以及本文讨论的三种协变量适应性随机化策略下，若在统计推断中未考虑协变量信息，相关方法往往会导致偏保守的一类错误控制，进而降低检验效能，因此不建议采用该类方法。

相较而言，采用分层分析方法或将协变量纳入回归模型的分析策略，在理论上可实现更高的统计效能。然而，这类方法的有效性前提在于：所有用于随机化的重要协变量 Z 必须被完整地纳入分析模型，且模型设定需正确。以分层分析为例，若未将所有分层因素 Z 所构成的联合分层结构完整地纳入分析，则该方法亦可能表现出偏保守的性质，降低推断效率。

此外，本文还总结并评述了若干在协变量适应性随机化设计中具有广泛适用性且能提升统计检验效能的“稳健”推断方法。通过实例分析与相应代码实现，本文展示了这些方法在实际应用中的具体表现，旨在为临床试验设计者和分析人员提供实践参考。

基于现有文献的梳理与比较，本文提出如下建议：对于连续型结局，推荐采用纳入随机化协变量的 ANHECOVA 方法；对于二分类结局，可考虑采用 ANHECOVA 或基于 AIPW 的线性矫正方法（LC）或联合矫正方法（JC）；对于生存结局数据，在满足比例风险假设且不存在稀疏分层的前提下，基于 Cox 比例风险模型的检验统计量 T_{CSL} 相较于 T_{CL} 具有更高的检验效能。然而，考虑到 T_{CL} 对模型假设依赖较少，且在统计效能方面已显著优于传统方法，综合评估后认为 T_{CL} 是更为稳妥的选择。

参考文献

[1] 国家药品监督管理局. 药物临床试验随机分配指导原则 (试行)[R]. 2022-01-04. (2022 年第 5 号通告)
[2] Friedman LM, Furberg CD, Demets DL, et al. Fundamentals of clinical trials [M]. 5th ed. Cham: Springer, 2015.
[3] Shao J, Yu X, Zhong B. A theory for testing hypotheses under covariate-adaptive randomization [J]. Biometrika, 2010, 97(2): 347-360.
[4] Ye T. Testing hypotheses under covariate-adaptive randomisation and additive models [J]. Stat Theory Relat Fields, 2018, 2(1): 96-101.
[5] Ye T, Shao J. Robust tests for treatment effect in survival analysis under covariate-adaptive randomization [J]. J R Stat Soc B, 2020, 82(5): 1301-1323.
[6] Bugni FA, Canay IA, Shaikh AM. Inference under covariate-adaptive randomization [J]. J Am Stat Assoc, 2018, 113(524): 1784-1796.
[7] Ye T, Shao J, Yi Y, et al. Toward better practice of covariate adjustment in analyzing randomized clinical trials [J]. J Am Stat Assoc, 2023, 118(544): 2370-2382.
[8] Shao J. Inference after covariate-adaptive randomisation: aspects of methodology and theory [J]. Stat Theory Relat Fields, 2021, 5(3): 172-186.
[9] Zelen M. The randomization and stratification of patients to clinical trials [J]. J Chronic Dis, 1974, 27(7): 365-375.
[10] Efron B. Forcing a sequential experiment to be balanced [J]. Biometrika, 1971, 58(3): 403-417.
[11] Kuznetsova OM, Johnson VP. Approaches to expanding the two-arm biased coin randomization to unequal allocation while preserving the unconditional allocation ratio [J]. Stat Med, 2017, 36(16): 2483-2498.
[12] Zhao W, Ramakrishnan V. Generalization of Wei's urn design to unequal allocations in sequential clinical trials [J]. Contemp Clin Trials Commun, 2016, 2: 75-79.
[13] Wei LJ. A class of designs for sequential clinical trials [J]. J Am Stat Assoc, 1977, 72(358): 382-386.
[14] Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial [J]. Biometrics, 1975, 31(1): 103-115.
[15] Taves DR. Minimization: a new method of assigning patients to treatment and control groups [J]. Clin Pharmacol Ther, 1974, 15(5): 443-453.
[16] Hu F, Zhang LX, He X. Efficient randomized-adaptive designs [J]. Ann Stat, 2009, 37(5A): 2543-2560.
[17] Ma W, Hu F, Zhang L. Testing hypotheses of covariate-adaptive randomized clinical trials [J]. J Am Stat Assoc, 2015, 110(510): 669-680.
[18] Ma W, Qin Y, Li Y, et al. Statistical inference for covariate-adaptive randomization procedures [J]. J Am Stat Assoc, 2020, 115(531): 1488-1497.
[19] Shao J, Yu X. Validity of tests under covariate-adaptive biased coin randomization and generalized linear models [J]. Biometrics, 2013, 69(4): 960-969.
[20] Wang B, Susukida R, Mojtabai R, et al. Model-robust inference for clinical trials that improve precision by stratified randomization and covariate adjustment [J]. J

Am Stat Assoc, 2023, 118(542): 1152-1163.
[21] Bannick MS, Shao J, Liu J, et al. A general form of covariate adjustment in clinical trials under covariate-adaptive randomization [J]. Biometrika, 2025, 112(3): asaf029.
[22] Ye T, Shao J, Yi Y. Covariate-adjusted log-rank test: guaranteed efficiency gain and universal applicability [J]. Biometrika, 2024, 111(2): 691-705.

An introduction to statistical analysis methods for clinical trials using covariate-adaptive randomization

YANG Yuxuan[1], ZHU Siyu[1], YIN Minggang[1], PAN Jianhong[2], DUAN Chongyang[1]

[1] State Key Laboratory of Multi-organ Injury Prevention and Treatment, Southern Medical University, Guangzhou 510515, Guangdong, China; [2] Center for Drug Evaluation, National Medical Products Administration, Beijing 100022, China

ABSTRACT Covariate-adaptive randomization is increasingly used in drug clinical trials because it helps balance prognostic factors that may affect trial outcomes between treatment groups. When performing statistical analyses of such trials, researchers usually consider it necessary to adjust for the covariates used in randomization. Therefore, regression models and stratified analyses are commonly employed for statistical inferences. However, no systematic studies have been conducted to discuss the rationality and applicability of these methods. To address this gap, the present study reviews the existing literature, examines the adaptability of traditional statistical inference methods for trials using covariate-adaptive randomization, and introduces the "robust" statistical inference methods proposed in recent years. In addition, to facilitate practical implementation by clinical researchers, example code is provided to demonstrate the application of these methods.

KEYWORDS adaptive-randomization; stratified analysis; covariate-adjusted; robust analysis method; software operation

附录 R 代码及复现说明

本附录提供正文实例分析所用的 R 代码，分别对应连续型结局、二分类结局和生存结局的分析。连续型结局代码用于比较 ANOVA、ANCOVA 和 ANHECOVA 方法；二分类结局代码用于比较 AIPW、线性校正（LC）和联合校正（JC）方法；生存结局代码用于比较传统 Log-rank 检验、分层 Log-rank 检验、Score 检验、稳健 Score 检验及协变量调整的 Log-rank 检验。相应分析结果分别见正文表 2、表 3 和表 4。

连续型结局和二分类结局分析使用 RobinCar 包内置模拟数据集 data_sim。生存结局分析使用 data_gen() 函数生成模拟数据；为保证该部分模拟结果可复现，代码中设置了随机种子。

运行前请加载 RobinCar 和 survival 包。不同版本的 RobinCar 包可能在函数参数或数据对象上存在差异，建议记录所使用的包版本。

A.1 加载 R 包

```
# 如未安装 RobinCar，请先运行：
# install.packages("RobinCar")

library(RobinCar)
library(survival)
#固定随机数种子
set.seed(2024)

# RobinCar 包相关函数的详细说明可参考：
#https://marlenabannick.com/RobinCar/articles/Calibration-for-GLM-Models.html
```

A.2 连续型结局分析：ANOVA、ANCOVA 和 ANHECOVA

以下代码使用 RobinCar 包中的模拟数据，对连续型结局变量 y 进行分析。治疗分组变量为 A；z1 和 z2 为随机化分层变量；x1 和 x3 为纳入协变量调整的变量。各方法的组间比较结果可用于生成正文表 2。

```
data <- RobinCar:::data_sim
data$A <- as.factor(data$A)
data$extra_cov <- rnorm(nrow(data))
# ANOVA
fit.anova <- robincar_linear(df = data,
                             response_col="y",
                             treat_col="A",
                            car_strata_cols=c("z1","z2"),
                             car_scheme="biased-coin",
                             adj_method="ANOVA")
fit.anova
robincar_contrast(fit.anova, contrast_h = "diff")
# ANCOVA
fit.ancova <- robincar_linear(df = data,
                             response_col="y",
                             treat_col="A",
                           car_strata_cols=c("z1", "z2"),
                            covariate_cols=c("x1", "x3"),
                             car_scheme="biased-coin",
                             adj_method="ANCOVA")
fit.ancova
robincar_contrast(fit.ancova, contrast_h = "diff")
# ANHECOVA
fit.anhecova <- robincar_linear(df = data,
                             response_col="y",
                             treat_col="A",
                           car_strata_cols=c("z1", "z2"),
                covariate_cols=c("z1","z2","x1","x3"),
                             car_scheme="biased-coin",
                             adj_method="ANHECOVA")
fit.anhecova
robincar_contrast(fit.anhecova, contrast_h = "diff")
```

A.3 二分类结局分析：AIPW、LC 和 JC

以下代码将连续型结局 y 按临界值 2 转换为二分类结局 y_bin，并比较 AIPW、LC 和 JC 方法。其中，joint = FALSE 表示线性校正，joint = TRUE 表示联合校正；extra_cov 为联合校正中额外纳入的协变量。分析结果可用于生成正文表 3。

```
# 数据载入
data <- RobinCar:::data_sim
data$A <- as.factor(data$A)
data$y_bin <- ifelse(data$y > 2, 1, 0)
data$extra_cov <- rnorm(nrow(data))
# ANOVA
fit.anova <- robincar_linear(df = data,
                             response_col="y_bin",
                             treat_col="A",
                            car_strata_cols=c("z1","z2"),
                             car_scheme="biased-coin",
                             adj_method="ANOVA")
fit.anova
robincar_contrast(fit.anova, contrast_h = "diff")
# ANCOVA
fit.ancova <- robincar_linear(df = data,
                             response_col="y_bin",
                             treat_col="A",
                           car_strata_cols=c("z1","z2"),
                            covariate_cols=c("x1","x3"),
                             car_scheme="biased-coin",
                             adj_method="ANCOVA")
fit.ancova
robincar_contrast(fit.ancova, contrast_h = "diff")
# ANHECOVA
fit.anhecova <- robincar_linear(df = data,
                             response_col="y_bin",
                             treat_col="A",
                           car_strata_cols=c("z1","z2"),
                covariate_cols=c("z1","z2","x1","x3"),
                             car_scheme="biased-coin",
                             adj_method="ANHECOVA")
fit.anhecova
robincar_contrast(fit.anhecova, contrast_h = "diff")
# AIPW
fit.aipw <- robincar_glm(
  df = data,
  response_col = "y_bin",
  treat_col = "A",
  car_strata_cols = c("z1"),
  car_scheme = "biased-coin",
```

```
 g_family = binomial(link = "logit"),
 g_accuracy = 7,
 formula = "y_bin ~ A + x3 + z1")
fit.aipw$result
robincar_contrast(fit.aipw, contrast_h = "diff")
# LC
fit.lc <- robincar_calibrate(
 result = fit.aipw,
 joint=FALSE)
fit.lc$result
robincar_contrast(fit.lc, contrast_h = "diff")
# JC
fit.jc <- robincar_calibrate(
 result=fit.aipw,
 joint=TRUE,
 add_x = c("extra_cov"))
fit.jc$result
robincar_contrast(fit.jc, contrast_h = "diff")
```

A.4 生存结局分析: 传统方法与协变量调整方法

以下代码采用 data_gen() 函数生成生存数据, 并比较传统 Log-rank 检验、分层 Log-rank 检验、Cox Score 检验、稳健 Cox Score 检验、协变量调整 Log-rank 检验和协变量调整分层 Log-rank 检验。分析结果可用于生成正文表 4。

```
# 生成数据
data <- data_gen(n = 200, theta = 0, randomization
= "CABC", p_trt = 0.5, case = "case1")
# classical log-rank test
logrank<-survdiff(Surv(t, delta) ~ I1, data = data)
print(logrank)

# stratified log-rank test
stratified.logrank <- coxph(Surv(t, delta) ~ I1 + stra-
ta(car_strata1, car_strata2, car_strata3, car_strata4,
car_strata5, car_strata6), data = data)
print(stratified.logrank)

# cox score test
score<-coxph(Surv(t, delta) ~ I1, data = data)
summary(score)

# Cox robust score test
cox <- robincar_coxscore(df = data,
                         treat_col = c("I1"),
                         response_col = "t",
                         event_col = "delta",
                          car_strata_cols = c("car_stra-
ta1", "car_strata2", "car_strata3", "car_strata4",
"car_strata5", "car_strata6"),
                         covariate_cols = c("model_Z1",
"model_Z21", "model_Z22"),
                         car_scheme = "biased-coin",
                         ref_arm = 0)
print(cox)

# covariate-adjusted logrank test
cl<-robincar_logrank(df = data,
                         treat_col = c("I1"),
                         response_col = "t",
                         event_col = "delta",
                          car_strata_cols = c("car_stra-
ta1", "car_strata2", "car_strata3", "car_strata4",
"car_strata5", "car_strata6"),
                         covariate_cols = c("model_Z1",
"model_Z21", "model_Z22"),
                         car_scheme = "biased-coin",
                         adj_method = "CL",
                         ref_arm = 0)
print(cl)

# Covariate-adjusted stratified logrank test
csl<-robincar_logrank(df = data,
                         treat_col = c("I1"),
                         response_col = "t",
                         event_col = "delta",
                          car_strata_cols = c("car_stra-
ta1", "car_strata2", "car_strata3", "car_strata4",
"car_strata5", "car_strata6"),
                         covariate_cols = c("model_Z1",
"model_Z21", "model_Z22"),
                         car_scheme = "biased-coin",
                         adj_method = "CSL",
                         ref_arm = 0)
print(csl)
```