多区域临床试验中协变量不匹配情况下区域特异性治疗效应推断的选择性信息借用

Chenxi Li[1], Ke Zhu[1,2],Shu Yang[2] 和 Xiaofei Wang[1], *

[1]杜克大学生物统计学与生物信息学系，美国北卡罗来纳州达勒姆，邮编 27710。[2]北卡罗来纳州立大学统计学系，美国北卡罗来纳州罗利，邮编 NC27695。*电子邮件：xiaofei.wang@duke.edu

arXiv:2602.02703v1 [stat.ME] 2 Feb 2026

SUMMARY: 多区域临床试验（MRCTs）是全球药物开发的核心，可用于评估多样化人群中的治疗效应。一项关键挑战是，当目标区域规模较小，且其基线协变量或未测量因素与辅助区域存在差异时，如何对区域特异性估计目标进行有效且高效的推断。我们采用基于估计目标的框架，聚焦于预先指定目标区域的区域特异性平均治疗效应（RSATE），该效应与地方监管决策直接相关。跨区域差异可能导致协变量偏移、协变量不匹配和结局漂移，从而使信息借用产生偏倚，并导致RSATE推断无效。为解决这些问题，我们提出了一个结合选择性信息借用的统一因果推断框架。首先，我们引入一种逆方差加权估计量，将“样本量小、协变量丰富”的仅目标估计量与“样本量大、协变量有限”的全量借用双重稳健估计量相结合，从而在不存在结局漂移时最大限度地提高效率。其次，为适应结局漂移，我们应用保形预测评估患者层面的可比性，并自适应地选择用于借用的辅助区域患者。第三，为确保严格的有限样本推断，我们采用条件随机化检验，实现精确、无模型且考虑选择的 I 类错误控制。模拟研究表明，与不借用和全量借用方法相比，所提出的估计量提高了效率，使均方误差降低 10–50%，检验功效更高，同时在多种情景下保持有效推断。对 POWER试验的应用进一步表明，该方法提高了RSATE估计的精度。

KEY WORDS: 因果推断；保形推断；随机化检验；可迁移性；I类错误控制

1. 引言

多区域临床试验（MRCT）已成为全球药物开发中的重要策略，使申办方能够评估不同多样化人群中的治疗有效性和安全性。通过招募来自多个地理位置分散区域的患者，MRCT能够提高试验效率、加快监管申报，并促进全球患者更快获得新疗法（Chen 等，2009，2010；Bean 等，2021，2023；Zhuang 等，2024；Hua 等，2024；Bean 等，2024；Robertson 等，2025；Dette 等，2025；Wan 等，2025；Alene 等，2025）。除了估计活性治疗的全局效应外，MRCT的另一主要目标是量化区域特异性治疗效应，以支持当地注册（ICH，1998年；Hung 等，2010年）。这对于监管审评尤为重要，因为当治疗效应在参与试验的国家之间存在差异时，地方卫生当局日益强调需要提供区域特异性证据，以证明批准决定的合理性（ICH，2017年；Guo 等，2016）。当目标区域的患者招募困难或区域亚人群规模较小时，就会面临一项关键挑战。样本量有限时，仅使用区域特异性数据估计治疗效应可能效率低下，因此有必要借鉴其他区域的信息，以改善对目标人群的推断。从正式的估计目标角度来看，这些问题自然可表述为目标区域人群中的区域特异性平均治疗效应，从而将试验设计和分析与地方监管决策问题明确联系起来。

然而，在跨区域借用信息时，必须考虑区域异质性，以避免有偏信息借用和无效推断。ICH(1998年) 承认，由于种族或遗传背景等固有差异，一个区域的数据可能无法直接外推至另一区域；ICH (2017年) 进一步建议预先设定借用或收缩策略，并对基线预后因素进行协变量调整。在实践中，这些考量带来了三项主要挑战。首先，

基线协变量偏移和不匹配十分常见。由于医疗基础设施、数据收集方案或伦理约束的差异，各区域的基线变量可用性或定义可能有所不同，因此各区域的协变量分布往往存在显著差异，有时协变量集甚至没有重叠。面对这类差异，直接借用信息会引入偏倚，而将分析限定于共享协变量则可能降低效率，尤其是在目标区域的其他地区特异性协变量具有较强预测能力时。其次，即使已测量协变量均衡，临床实践模式等未测量因素仍可能因区域而异，从而在目标区域与辅助区域之间造成隐藏偏倚。这个问题通常称为结局漂移。第三，有效推断仍然困难。目标区域样本量有限，可能会削弱大样本近似的可靠性；依赖倾向评分或结局模型的协变量偏移调整，在两个模型均设定错误时可能失效；而针对结局漂移设计的动态借用程序，若未恰当纳入选择不确定性，也可能导致无效推断。

为应对信息借用过程中出现的挑战，数据整合文献中已开发出广泛的方法（Colnet 等，2024）。针对数据来源之间的协变量偏移，已有研究提出倾向评分匹配与加权、结局回归以及双稳健估计量（Valancius 等，2024）。协变量不匹配问题日益受到关注（Han 等，2023、2024；Zeng 等，2025；Li 等，2025；Williams 等，2025），在迁移学习文献中也称为块缺失（Chang 等，2024；Xu 和 Qu，2025）。针对结局漂移，贝叶斯动态借用方法会根据来源信息与目标数据的相似程度，自适应地降低来源信息的权重；这类方法包括信息性先验（Chen 和 Ibrahim，2000；Hobbs 等，2011；Schmidli 等，2014；Yang 等，2023；Kwiatkowski 等，2024；Alt 等，2024）和贝叶斯层次模型（Thall 等，2003；Kaizer 等，2018）。近期，研究者提出了若干对结局漂移具有稳健性的频率学派策略，包括预后

调整或数字孪生（Schuler 等，2022；Liu 等，2025）、先检验后合并（Viele 等，2014；Yang 等，2023；Gao 和 Yang，2023；Dang 等，2025）、选择性借用（Gao 等，2025a、b；Zhu 等，2025；Liu 等，2025）、偏差校正（Stuart 和 Rubin，2008；Wu 和 Yang，2022；Yang 等，2024；Mao 等，2025；Ye 等，2025；van der Laan 等，2024）、聚合组合（Chen 等，2025；Cheng 和 Cai，2021；Rosenman 等，2023），以及利用辅助变量的方法（Guo 等，2022；Dang 等，2025）。为确保借用后的推断有效，贝叶斯程序通常需要额外校准以控制I类错误，而频率学派方法则有多种选择：（i）假设不存在结局漂移时的渐近正态性（Li 等，2023；Valancius 等，2024），或在预言机选择无偏来源的情况下（Gao 等，2025a），或在对偏倚进行预言机校正的情况下（Yang 等，2024；van der Laan 等，2024）；（ii）能够考虑选择不确定性的渐近非正态分布（Yang 等，2023；Dang 等，2025）；或（iii）纳入选择不确定性的随机化检验或置换检验（Zhu 等，2025；Liu 等，2025；Ren 等，2025）。在多区域临床试验场景中，Robertson 等（2025）提出了一种不借用信息的仅目标估计量，以及一种在不存在结局漂移时借用全部信息的双重稳健估计量，并通过对区域与结局关联进行预检验，在二者之间作出选择；这实质上遵循了先检验后合并策略。他们的推断依赖渐近正态性，且未考虑预检验引入的不确定性。

在这些进展的基础上，并将区域特异性平均治疗效应（RSATE）作为目标区域的正式估计目标，我们在本文中做出三项贡献。首先，我们提出一种逆方差加权（IVW）方法，将仅目标估计量（小样本、丰富的协变量）与全量借用双重稳健估计量（大样本、有限协变量）相结合，以在无结局漂移假设下最大化效率。其次，认识到将辅助区域数据作为整体使用时，结局漂移假设可能不成立，我们评估个体可比性，对象为

每位患者，并使用保形预测（Vovk等，2005年；Lei等，2018年；AngelopoulosandBates，2023年），自适应地确定一个阈值，以选择满足无结局漂移假设的患者子集，然后将逆方差加权估计量应用于目标区域数据以及这一已选择的辅助区域数据子集。第三，为便于进行有效推断，我们采用条件随机化检验，以提供有限样本精确、无模型且选择感知的推断（Fisher，1935年；Zhengand Zelen，2008年；Zhang and Zhao，2023年）。综合来看，这些组成部分构成了一个稳健且高效的基于估计目标的框架，用于在选择性信息借用下推断区域特异性平均处理效应，从而应对协变量偏移与不匹配、结局漂移以及有限样本 I 类错误控制，并与多区域临床试验中的本地监管评估直接契合。

本文结构如下。第2 节介绍多区域临床试验的一般因果推断框架。第3节介绍协变量偏移和不匹配情况下的信息借用方法。第4 节介绍所提 CSB 方法。第5 节介绍一种随机化推断框架，该框架在严格提高检验功效的同时控制 I 类错误率。第6 节展示模拟研究的设计与结果。第7 节将所提方法应用于 POWER 试验数据。最后，我们在第8节进行讨论。

2. 预备知识

2.1 目标区域、估计目标与数据结构

我们以POWER试验作为启发性示例，引入符号和潜在结果框架。该试验从多个地理上不同的地区招募患者，包括北美、南美洲和欧洲，总样本量为 n。我们的主要兴趣并非针对所有地区的合并总体，而是预先指定的目标区域，例如北美，这与支持本地监管决策的目标一致。因此，我们将研究样本划分为来自目标区域的 $n_{\mathcal{R}}$ 名患者（ $R=1$ ）和来自辅助区域的 $n_{\mathcal{E}}$ 名患者（ $R=0$），其中R 表示样本指示变量。辅助区域包括所有非目标区域，例如

南美洲和欧洲。我们不再进一步区分辅助区域，因为异质性将在第4节中在个体患者层面进行处理。

我们关注的治疗比较是活性治疗 $A = 1$ (enobosarm) 与对照组 $A = 0$ (安慰剂) 之间的比较。令 $Y(a)$ 表示接受治疗a时的潜在结果，定义为第84天瘦体重的百分比变化。我们关注的估计目标是区域特异性平均治疗效应（RSATE），定义为 $\tau = \theta_1 - \theta_0$，其中$\theta_a = \mathbb{E}[Y(a) \mid R = 1]$ 表示目标区域人群中接受治疗 a 时潜在结果的均值。在POWER试验中， n患者被随机分配至样本量为 n_1 的治疗组 $A = 1$，或样本量为 n_0 的对照组 $A = 0$；其中A 为治疗指标，观测结局记为 Y。

所有区域均收集一组共同的基线协变量，例如年龄、性别和 ECOG 体能状态，并记为 X。一般而言，目标区域可能与不同辅助区域共享不同的协变量集合；我们在补充材料的附录 F 中探讨这种情况。此外，由于数据收集、定义或测量实践存在差异，目标区域还可能包含辅助区域中无法获取或无法直接比较的区域特异性基线协变量 U。例如，北美区域记录了体重下降标志，但辅助区域没有记录。因此，观测数据结构为 $\{R_i, X_i, U_i, A_i, Y_i\}_{i=1}^n$，其中 U_i 对于 $R_i = 0$个体而言缺失。表 1 和图 1 总结了观测数据结构。

[Table 1 about here.]

[Figure 1 about here.]

2.2 仅目标区域识别与估计

我们首先介绍以下识别假设（Robertson 等人，2025）.

Assumption 1: (i)（一致性） $Y = AY(1) + (1 - A)Y(0)$。(ii)（无混杂性）$Y(a) \perp A \mid X, R = 1$ 对于 $a \in \{0, 1\}$。(iii)（正值性） $0 < \mathbb{P}(A = 1 \mid X, R = 1) < 1$。

假设 n 1这通常可通过多区域临床试验的设计来满足。第 (i) 部分要求个体之间不存在干扰，且不存在会导致同一个体产生不同结局的多种处理版本。

第 (ii) 部分假设在目标区域内，给定基线协变量后不存在未测量混杂。在多区域临床试验中，治疗分配通常在设计上与协变量和地区相互独立；采用区域分层随机化时，治疗分配则在各区域内与协变量相互独立。这里，我们采用一个较弱的条件，仅要求在目标区域内给定基线协变量后，潜在结果与治疗分配相互独立。第 (iii) 部分要求目标区域内每个个体在给定协变量后，接受任一治疗水平的概率均大于零；这通常在多区域临床试验中成立。

在假设 1 下，仅使用目标区域的数据即可识别区域特异性平均治疗效应 τ。令 $e_{a|1}(x) = \mathbb{P}(A = a \mid R = 1, X = x)$ 表示目标区域内的治疗倾向评分；在多区域临床试验中，该评分由设计确定。令 $\mu_a(x) = \mathbb{E}[Y(a) \mid X = x, R = 1]$ for$a \in \{0, 1\}$ 表示目标区域内基于共同基线协变量 X 的条件平均结局函数。仅依赖目标区域数据并对 X 进行调整的 τ 协变量调整估计量为 $\hat{\tau}^{\text{NB-Xonly}} = \hat{\theta}_1^{\text{NB-Xonly}} - \hat{\theta}_0^{\text{NB-Xonly}}$，其中

$$\hat{\theta}_a^{\text{NB-Xonly}} = \frac{1}{n_{\mathcal{R}}} \sum_{i=1}^{n} \left[R_i \hat{\mu}_a^{\text{NB}}(X_i) + \frac{\mathbb{I}(R_i = 1, A_i = a)}{e_{a|1}(X_i)} \{Y_i - \hat{\mu}_a^{\text{NB}}(X_i)\} \right].$$

这里，$\hat{\mu}_a^{\text{NB}}(x)$ 是使用目标区域中满足 $R = 1$ 的观测值拟合的 $\mu_a(x)$ 结局回归模型。上标“NB”表示“不借用”，因为未纳入辅助区域的任何信息。

估计量 $\hat{\tau}^{\text{NB-仅 X}}$ 未利用目标区域特异性协变量 U，而这些协变量通常具有很强的预后作用，调整后可提高效率。令 $\mu_a(x, u) = \mathbb{E}[Y(a) \mid X = x, U = u, R = 1]$ 表示基于目标区域所有预后变量的条件平均结局函数。这里稍微滥用了符号，使用相同的符号 μ_a，并通过自变量加以区分。一个同时纳入 X 和 U 的协变量调整估计量为 $\hat{\tau}^{\text{NB-AllCov}} = \hat{\theta}_1^{\text{NB-AllCov}} - \hat{\theta}_0^{\text{NB-AllCov}}$,

其中

$$\hat{\theta}_a^{\text{NB-AllCov}} = \frac{1}{n_{\mathcal{R}}} \sum_{i=1}^{n} \left[R_i \hat{\mu}_a^{\text{NB}}(X_i, U_i) + \frac{\mathbb{I}(R_i = 1, A_i = a)}{e_{a|1}(X_i)} \{Y_i - \hat{\mu}_a^{\text{NB}}(X_i, U_i)\} \right].$$

此处，$\hat{}\mu_a^{\text{NB}}$（x，u）是仅使用目标区域数据、针对 μ_a（x，u）建立的拟合的结局模型，可采用自适应套索（Zou，2006）等数据自适应方法，或借助领域知识，从（X，U）中选择强预后变量。

$\hat{}\tau^{\text{NB-仅 X}}$ 与 $\hat{}\tau^{\text{NB-AllCov}}$ 均利用治疗倾向评分，因而对结局模型错误设定具有稳健性。估计量 $\hat{}\tau^{\text{NB-AllCov}}$ 通常效率更高，因为它对目标区域中的全部预后变量进行了调整。在标准正则条件下，两个估计量均相合且渐近正态，其方差估计量具有闭式形式；更多细节见附录 A.1。然而，这两个估计量均未从辅助区域借用信息，这促使下一节引入借用策略，以进一步提高效率。

2.3 辅助区域整合的识别与估计

为了实现从辅助区域的信息借用，我们引入以下假设（Robertson 等人，2025），该假设将在第 4 节中放宽。

Assumption 2（条件均值可交换性）：$\mathbb{E}[Y$（a）$\mid R = 0,\ X] = \mathbb{E}[Y$（$a$）$\mid R = 1,\ X]$ 对于 $a \in \{0, 1\}$。

假设 2 允许不同区域的结局分布存在差异，但要求目标区域与辅助区域之间的任何系统性差异都能由共同观测协变量完全解释 X。换言之，在对 X进行条件化后，假设各区域的条件均值潜在结局保持不变。如果存在未测量协变量，例如 U，同时影响区域归属和结局，则该假设可能不成立。我们将假设 2 视为一种理想化条件，并在此假设下制定信息借用程序。在第 4 节，我们放宽这一要求——

允许只有一部分辅助区域观测满足条件均值可交换性，并有选择地从这些观测中借用信息。

在假设2下，区域特异性平均治疗效应 τ 可利用目标区域和辅助区域的数据进行识别。令 $e_a(x) = \mathbb{P}(A = a \mid X = x)$表示所有区域的治疗倾向评分，该评分在多区域临床试验中由设计确定。令 $\pi(x) = \mathbb{P}(R = 1 \mid X = x)$表示抽样倾向评分，用于调整目标区域与辅助区域之间的协变量偏移。该值通常未知，可利用所有区域的数据拟合抽样模型 $R \sim X$ 进行估计，拟合值记为$\hat{\pi}^{\mathrm{FB}}(x)$。“FB”上标表示“完全借用”，因为根据假设2，纳入了所有辅助区域观测。

借用所有区域信息并调整 X 的协变量偏移的 τ 双重稳健估计量为$\hat{\tau}^{\text{FB-Xonly}} = \hat{\theta}_1^{\text{FB-Xonly}} - \hat{\theta}_0^{\text{FB-Xonly}}$，其中

$$\hat{\theta}_a^{\text{FB-Xonly}} = \frac{1}{n_{\mathcal{R}}} \sum_{i=1}^{n} \left[R_i \, \hat{\mu}_a^{\mathrm{FB}}(X_i) + \hat{\pi}^{\mathrm{FB}}(X_i) \frac{\mathbb{I}(A_i = a)}{e_a(X_i)} \{ Y_i - \hat{\mu}_a^{\mathrm{FB}}(X_i) \} \right].$$

此处，$\hat{\mu}_a^{\mathrm{FB}}(x)$ 是利用目标区域和辅助区域中的观测数据构建的、针对 $\mu_a(x)$ 的拟合结局回归模型。

在假设 2 下，估计量 $\hat{\theta}_a^{\text{FB-Xonly}}$具有双重稳健性：只要结局回归模型或抽样倾向评分模型中任一模型设定正确，它便是一致的；当两个模型的设定均正确时，它具有半参数有效性。在正则条件下，$\hat{\theta}_a^{\text{FB-Xonly}}$渐近正态，且其方差估计量有闭式表达；更多详情见补充材料中的附录 A.2。

Robertson 等人（2025）建议在假设 2 成立时使用 $\hat{\tau}^{\text{FB-Xonly}}$，在假设 2 不成立时使用 $\hat{\tau}^{\text{NB-AllCov}}$。他们的分析未考虑协变量不匹配问题，而且在其设定中没有可用的额外目标区域特异性协变量 U。因此，在他们的框架中，$\hat{\tau}_{\text{NB-AllCov}}$简化为 $\hat{\tau}^{\text{NB-Xonly}}$。尽管这些建议在其假设下是合理的，但在有额外协变量信息可用、且可交换性可能仅部分成立的情形下，仍可进一步改进：

这些建议在其假设下是合理的，但在有额外协变量信息可用、且可交换性可能仅部分成立的情形下，二者仍可进一步改进：

- 当假设2 成立时，目标区域中存在额外协变量 U 意味着 $\hat{\theta}^{\text{FB-Xonly}a}$ 的效率可能低于 $\hat{\theta}^{\text{NB-AllCov}a}$，因为前者忽略了其中包含的预测信息 U。我们在第3 节通过开发能够纳入 U 同时仍从辅助区域借用信息的估计量，来解决这一局限。

- 当假设2 在整体上不成立，但在辅助区域的某个子群体中成立时，舍弃所有辅助区域信息、仅依赖 $\hat{\theta}^{\text{NB-AllCov}a}$ 并非最优做法。相反，我们有选择地借用满足第4节中可交换性条件的辅助观测信息。

3. 协变量不匹配下的信息借用

在假设2下，我们进一步利用目标区域特异性协变量 $\theta_a^{\text{FB-Xonly}}$所带来的额外预测能力，以改进估计量 $\hat{}\, U$。对于满足$R_i = 1$条件的目标区域患者，可以通过两种方式预测其结局 Y_i: (i) $\hat{}\mu_a^{\text{NB}}(X_i,\ U_i)$，该方法利用更丰富的协变量信息，但基于较少的观测进行估计；(ii) $\hat{}\mu_a^{\text{FB}}(X_i)$，该方法使用较有限的协变量集，但基于更大的合并样本进行估计。我们的主要思路是使用逆方差加权（IVW）结合这两种预测。

设 $\hat{}\mu_a^{\text{NB}}(X_i, U_i)$ 和 $\hat{}\mu_a^{\text{FB}}(X_i)$ 的经验预测误差为

$$v_{\text{NB}} = \frac{\sum_{i=1}^n \mathbb{I}(A_i = a, R_i = 1)\{Y_i - \hat{\mu}_a^{\text{NB}}(X_i, U_i)\}^2}{\sum_{i=1}^n \mathbb{I}(A_i = a, R_i = 1)}, \quad v_{\text{FB}} = \frac{\sum_{i=1}^n \mathbb{I}(A_i = a)\{Y_i - \hat{\mu}_a^{\text{FB}}(X_i)\}^2}{\sum_{i=1}^n \mathbb{I}(A_i = a)}.$$

逆方差加权预测定义为

$$\hat{Y}_i^{\text{FB}} = \begin{cases} \dfrac{v_{\text{FB}}}{v_{\text{NB}} + v_{\text{FB}}} \hat{\mu}_a^{\text{NB}}(X_i, U_i) + \dfrac{v_{\text{NB}}}{v_{\text{NB}} + v_{\text{FB}}} \hat{\mu}_a^{\text{FB}}(X_i), & R_i = 1, \\ \hat{\mu}_a^{\text{FB}}(X_i), & R_i = 0. \end{cases}$$

由于目标区域特异性协变量 U 对于辅助区域观测不可用，因此对具有 $R_i = 0$ 的个体进行预测时，仅依赖于 $\hat{}\mu_a^{\text{FB}}(X_i)$。

在 $\hat{}\theta_a^{\text{FB-Xonly}}$ 中用 $\hat{}Y^{\text{FB}i}$ 替换 $\hat{}\mu_a^{\text{FB}}(X_i)$，得到 $\hat{}\tau^{\text{FB-IVW}} = \hat{}\theta_1^{\text{FB-IVW}} - \hat{}\theta_0^{\text{FB-IVW}}$，其中

$$\hat{\theta}_a^{\text{FB-IVW}} = \frac{1}{n_{\mathcal{R}}} \sum_{i=1}^{n} \left[R_i \hat{Y}_i^{\text{FB}} + \hat{\pi}^{\text{FB}}(X_i) \frac{\mathbb{I}(A_i = a)}{e_a(X_i)} \{Y_i - \hat{Y}_i^{\text{FB}}\} \right]. \tag{1}$$

在假设 1 和 2下，估计量 $\hat{}\theta_a^{\text{FB-IVW}}$具备 $\hat{}\ \theta_a^{\text{FB-Xonly}}$的稳健性和效率特性，同时进一步利用 U 中包含的预后信息，并通过逆方差加权实现这一点。这种构造特别适用于各地区存在协变量不匹配的情形，因为它既避免舍弃辅助区域观测，又能有效纳入额外的目标区域特异性协变量。从另一种角度看，(1)可视为一个估计量加上一个零值的一致估计量；这一思路在概念上与 Yang 和 Ding (2020)的增广框架相关；更多细节见补充材料中的附录A.3。

4. 结局漂移下的保形选择性借用

虽然假设 2为将辅助区域数据纳入目标区域治疗效应估计提供了依据，但由于未测量的地区差异（例如医疗卫生系统或数据收集方案的差异），该假设在实践中可能不成立。因此，当各地区存在条件结局不相容时，依赖假设 2 的完全借用方法可能会导致估计偏倚和无效推断。

幸运的是，由于可获得目标区域的数据作为基准方法，因此可以利用观测数据对假设 2 是否违背进行实证评估（Robertson 等人，2025）。当证据表明假设2 被违背时，一种保守策略是舍弃所有辅助区域数据，仅依赖目标区域信息。然而，如果假设2 对部分辅助区域观测成立（启发性示例中正是如此，见图4），这种做法可能造成显著的效率损失。

这促使我们致力于有选择地识别与目标区域可交换的辅助区域观测，并仅从这些观测中借用信息。我们分两步实现这一目标。首先，我们利用保形p值评估辅助区域观测在

个体层面上的可交换性。其次，我们通过最小化一类 RSATE 估计量的均方误差（MSE）来确定选择阈值，从而根据辅助区域观测对应的保形p值作出整体借用决策。由于选择阈值是数据自适应的，我们在第 5 节建议采用条件随机化推断，在每次对目标区域重新随机化时重新执行选择程序，以保持精确的 I 类错误控制。我们构造类似于（1）的治疗效应估计量，仅使用来自目标区域和已选择的辅助区域观测的数据。我们将这一程序称为保形选择性借用（CSB），并在以下小节中详细介绍其方法。

4.1 共形 p 值

在本小节中，我们介绍如何为单个辅助区域观测值 (X_j, Y_j) 计算保形 p 值，其中 $R_j = 0$ 和 $A_j = a$。目标是评估该观测值相对于目标区域中相应治疗组的可交换性，即 $\{(X_i, Y_i)\}_{i\in\mathcal{R}_a}$，其中 $\mathcal{R}_a = \{i : R_i = 1, A_i = a\}$ 表示接受处理 a 的目标区域患者的索引集。步骤如下。

1. 数据拆分：将 $\mathcal{R}_a$ 随机划分为 K 个互不重叠的折，$\mathcal{R}_a = \bigcup_{k=1}^{K} \mathcal{R}_a^k$。

2. 训练：对于每一折 k，使用其余的$K-1$ 折数据拟合回归函数 $\hat{}\mu_a^{-k}(X)$，并排除 $\mathcal{R}_a^k$。

3. 共形分数：对于辅助区域观测值 (X_j, Y_j)，计算其相对于每个目标区域观测值的共形分数，具体如下

$$s_j^{(i)} = \left|Y_j - \hat{\mu}_a^{-k(i)}(X_j)\right|, \qquad i \in \mathcal{R}_a,$$

其中，$k(i) \in \{1, \ldots, K\}$ 表示满足 $i \in \mathcal{R}_a^{k(i)}$ 的折叠索引。

4. 校准：对于每个目标区域观测值 (X_i, Y_i)，计算共形分数

$$s_i = \left|Y_i - \hat{\mu}_a^{-k(i)}(X_i)\right|, \qquad i \in \mathcal{R}_a.$$

5. CV+ p值：将（X_j, Y_j）的 CV+ 保形 p 值定义为

$$p_j = \frac{\sum_{i \in \mathcal{R}_a} \mathbb{I}\{s_i \geqslant s_j^{(i)}\} + 1}{|\mathcal{R}_a| + 1}.$$

直观来说，如果 (X_j, Y_j) 与目标区域观测 $\{(X_i, Y_i)\}_{i \in \mathcal{R}_a}$ 不可交换，那么相对于根据目标区域观测计算得到的 $\{s_i\}_{i \in \mathcal{R}_a}$，其共形分数 $\{s_j^{(i)}\}_{i \in \mathcal{R}_a}$ 往往会较大。因此，p值 p_j 会较小，这为反对可交换性提供了证据，并表明不应借用 (X_j, Y_j)。

Remark 1 （共形分数）：虽然我们以绝对残差作为共形分数进行说明，但上述流程更普遍地适用于任何用于量化 (X_j, Y_j) 与目标区域观测之间相似性的共形分数。例如，保形分位数回归（CQR）分数（Romano 等人，2019）基于拟合的条件分位数，而基于最近邻的分数可用于二元结局（Shafer 和 Vovk，2008）。因此，所提出的流程无模型且广泛适用于现代信息借用场景。

Remark 2 （分割保形 p 值）：使用分割保形方法可以获得一种更简单的保形 p 值（Papadopoulos 等人，2002），该方法依赖于将数据一次性划分为训练集和校准集。我们采用 CV+ 方法（Barber 等人，2021），因为目标区域样本 $\mathcal{R}_a$ 通常数量有限，而 CV+ 通过利用所有观测值进行训练和校准，能够更高效地使用可用数据。初步模拟结果表明，与分割保形 p 值相比，CV+ 能够提升有限样本选择性能。

4.2 MSE 引导的选择阈值

对上一小节中的程序进行应用，可得到所有辅助区域观测的保形p值。较大的 p_j 表示与目标区域的相容性更高，但需要设定选择阈值 $\gamma_a \in [0, 1]$，以确定要借用于估计 p_j 的辅助区域观测子集

。具体而言，对于治疗组 a，我们将已选择的集合定义为 $\mathcal{E}_a(\gamma_a) = \{j\colon R_j = 0, A_j = a, p_j \geqslant \gamma_a\}$.

一种可能的方法是将其表述为一个多重检验问题，并选择 γ_a 以控制错误发现率（Jin and Cand`es，2023年）。然而，由于目标区域的样本量有限，这种 γ_a 可能导致检验功效较低，难以检测出不相容的辅助区域观测，并因此使所得的 RSATE 估计量产生较大偏倚。相反，我们将保形 p 值视为相似性排序，并确定 γ_a，以最小化下文介绍的 RSATE 估计量的均方误差。这一视角与因果估计量的目标调优密切相关（Rothenh¨ ausler，2024；Zhu 等人，2025）。

给定阈值 γ_a，我们使用目标区域数据和已选择的辅助区域观测，构造一个类似于（1）的 RSATE 估计量，

$$\hat{\theta}_a^{\text{CSB-IVW}}(\gamma_a) = \frac{1}{n_{\mathcal{R}}} \sum_{i=1}^{n} \left[R_i \hat{Y}_i^{\text{CSB}} + \hat{\pi}^{\text{CSB}}(X_i) \frac{\mathbb{I}(A_i = a, p_i \geqslant \gamma_a)}{\hat{e}_a^{\text{CSB}}(X_i)} \{Y_i - \hat{Y}_i^{\text{CSB}}\} \right],$$

其中，对于所有满足 $R_i = 1$ 的目标区域观测，我们令 $p_i \equiv 1$ 。量 $\hat{}\pi^{\text{CSB}}$ 和 $\hat{}Y_i^{\text{CSB}}$ 的定义与 $\hat{}\pi^{\text{FB}}$ 和 $\hat{}Y^{\text{FB}i}$ 类似，不同之处在于，完整的辅助区域观测集 $\{(X_j, Y_j)\}_{j:R_j=0,\ A_j=a}$ 被选择性借用子集 $\{(X_j, Y_j)\}_{j\in\mathcal{E}_a\ (\gamma_a)}$ 所取代。此外，$\hat{}e_a^{\text{CSB}}(X_i)$ 表示条件概率 $\mathbb{P}[A_i = a,\ p_i \geqslant \gamma_a \mid X_i]$ 的一个估计量。为简化记号，我们省略这些量对 γ_a 的依赖。

这一公式表明，$\hat{}\,\theta_a^{\text{CSB-IVW}}(\gamma_a)$在完全借用估计量和不借用估计量之间进行插值。当 $\gamma_a = 0$, $\mathcal{E}_a(0) = \{j\colon\ R_j = 0, A_j = a\}$ 且$\hat{}\,\theta_a^{\text{CSB-IVW}}(0)$与完全借用估计量$\hat{}\,\theta_a^{\text{FB-IVW}}$一致。当 $\gamma_a = 1$, $\mathcal{E}_a(1) = \emptyset$ 且$\hat{}\,\theta_a^{\text{CSB-IVW}}(1)$退化为不借用估计量$\hat{}\,\theta_a^{\text{NB-AllCov}}$。对于中间值 $0 < \gamma_a < 1$，$\hat{}\theta_a^{\text{CSB-IVW}}(\gamma_a)$会有选择地借用来自兼容性最高的辅助区域观测的信息。

我们将均方误差 $\text{MSE}_a(\gamma_a)$ 分解为 $= [\mathbb{E}\{\hat{}\theta_a^{\text{CSB-IVW}}(\gamma_a)\} - \theta_a]^2 + \mathbb{V}\{\hat{}\theta_a^{\text{CSB-IVW}}(\gamma_a)\}$. 由于 θ_a 未知，我们使用一致的不借用估计量 $\hat{}\,\theta_a^{\text{NB-AllCov}}$作为基准来近似偏差平方，$[\mathbb{E}\{\hat{}\,\theta_a^{\text{CSB-IVW}}(\gamma_a)\} - \theta_a]^2 \approx \mathbb{E}\{\hat{}\theta_a^{\text{CSB-IVW}}(\gamma_a) - \hat{}\theta_a^{\text{NB-AllCov}}\}^2 - \mathbb{V}\{\hat{}\theta_a^{\text{CSB-IVW}}(\gamma_a) -$

$\hat{\theta}_a^{\text{NB-AllCov}}\}$. 方差项可通过非参数自助法近似，从而得到$\widehat{\qquad}$ $\widehat{\qquad}$
均方误差 $\text{MSE}_a(\gamma_a)$的估计值。我们在网格 $\gamma_a \in \mathcal{G} \subset [0,1]$ 上评估 $\text{MSE}_a(\gamma_a)$，并选择 $\hat{\gamma}_a = \arg\min_{\gamma_a \in \mathcal{G}} \widehat{\text{MSE}}_a(\gamma_a)$。最终估计量为 $\hat{\tau}^{\text{CSB-IVW}} = \hat{\theta}_1^{\text{CSB-IVW}}(\hat{\gamma}_1) - \hat{\theta}_0^{\text{CSB-IVW}}(\hat{\gamma}_0)$，该 MSE 引导的阈值选择策略是一项预先指定的数据自适应规则，旨在相较于固定选择 γ_a改善有限样本性能。由于 $\hat{\gamma}_a$ 是根据观测数据选择的，我们建议在假设检验中采用第 5 节的条件费希尔随机化检验，以考虑选择不确定性。由于 FB 和 NB 估计量均作为特例纳入其中，因此预期所得估计量的表现优于二者。详细算法见补充材料中的附录 B。

5. 用于 I 类错误率控制的 Fisher 随机化检验

获得 RSATE 估计量 $\hat{\tau}^{\text{CSB-IVW}}$后，在满足以下条件时，可以进行渐近推断：(i) 结局回归模型或倾向评分模型之一设定正确；(ii) 对满足假设 2 的辅助区域观测的选择是一致的。为进一步确保推断的有效性，我们还提出了一种有限样本精确、无模型且选择后有效的 Fisher 随机化检验（FRT），即使这些假设不成立，该检验也能控制 I 类错误率。具体步骤如下。

1. 目标区域中的尖锐零假设：在尖锐零假设 $H_0: Y_i(0) = Y_i(1)$ 对所有 $i \in \mathcal{R} = \{i: R_i = 1\}$ 成立时，可以利用观测结果 Y_i填补目标区域中所有缺失的潜在结果。

2. 检验统计量：指定一个检验统计量，并根据观测到的分配向量 $\boldsymbol{A} = (A_1, \ldots, A_n)$ 计算其观测值 $T(\boldsymbol{A})$。

3. 条件随机化：将辅助区域观测的处理分配固定为其观测值，即对所有满足 $R_i = 0$ 的 i，设定 $A_i^* = A_i$。按照原始随机化设计，对目标区域观测的处理分配 A_i^* 进行随机分配，从而得到重新分配后的向量 $\boldsymbol{A}^* = (A_1^*, \ldots, A_n^*)$。

4. p值：重复生成 $\boldsymbol{A}^*$，并计算相应的检验统计量 $T(\boldsymbol{A}^*)$。FRT p值定义为 $p^{\mathrm{FRT}} = \mathbb{P}_{\boldsymbol{A}^*}\{\,|T(\boldsymbol{A}^*)| \geqslant |T(\boldsymbol{A})|\,\}$，其中概率是相对于条件随机化分布计算的。

Remark 3 （检验统计量）：FRT对任何检验统计量的选择均有效。因此，本文所引入的任何估计量都可用作检验统计量，包括$\hat{\tau}^{\text{FB-IVW}}$ 即使在假设2 不成立时也是如此，因为FRT在精确零假设下能正确再现其有限样本分布。然而，当借用有偏倚的辅助区域观测时，相对于基于$\hat{\tau}^{\text{NB-AllCov}}$的FRT，该检验的检验功效可能会降低。因此，我们建议使用$\hat{\tau}^{\text{CSB-IVW}}$ 作为检验统计量，因为它有选择地借用相容的辅助区域信息，并可提高检验功效。一个关键点是，在计算$T(\boldsymbol{A}^*)$在随机化下时，必须允许CSB程序针对 $\boldsymbol{A}^*$ 的每一次实现重新选择辅助区域观测，以便正确考虑选择不确定性。

Remark 4 （蒙特卡洛近似）：在实践中，可通过 $\boldsymbol{A}^*$ 的蒙特卡洛抽样来近似随机化分布，而不是枚举所有可能的处理分配。

Remark 5 （单侧假设和基于随机化的区间）：上述讨论聚焦于双侧检验。对于单侧假设，例如 H_0： $Y_i(0) > Y_i(1)\ \forall i \in \mathcal{R}$，FRT 的 p 值定义为 $p^{\mathrm{FRT}} = \mathbb{P}_{\boldsymbol{A}^*}\{T(\boldsymbol{A}^*) \geqslant T(\boldsymbol{A})\}$.基于随机化的区间可通过对 FRT 求逆得到（Luo等人，2021；Zhu和Liu，2023年，2024）。

以下定理确立了 FRT 的有效性；其证明见补充材料中的附录 C。

Theorem 1: 在零假设 H_0 下，对于任意 $\alpha \in (0, 1)$，$\mathbb{P}_{\boldsymbol{A}}\left(p^{FRT} \leqslant \alpha \mid \boldsymbol{A}_{\mathcal{E}}\right) \leqslant \alpha$，其中 $\mathbb{P}_{\boldsymbol{A}}$ 表示在将潜在结果和协变量视为固定的情况下，相对于分配向量 $\boldsymbol{A}$ 的随机化分布的概率，而 $\boldsymbol{A}_{\mathcal{E}}$ 表示与辅助区域观测相对应的子分配向量。此外，

$$\mathbb{P}_{\boldsymbol{A}}\left(p^{FRT} \leqslant \alpha\right) = \sum_{\boldsymbol{A}_{\mathcal{E}}} \mathbb{P}_{\boldsymbol{A}}\left(p^{FRT} \leqslant \alpha \mid \boldsymbol{A}_{\mathcal{E}}\right) \mathbb{P}(\boldsymbol{A}_{\mathcal{E}}) \leqslant \alpha \sum_{\boldsymbol{A}_{\mathcal{E}}} \mathbb{P}(\boldsymbol{A}_{\mathcal{E}}) = \alpha.$$

6. 模拟研究

6.1 模拟设计

为评估所提方法的有限样本性能，并考察协变量的信息量如何影响性能，我们在两种情景下针对 X 的不同信号强度开展模拟研究：(i) $X = (X_1, X_2)$ 与 U 独立生成；(ii) X 与 U 联合生成，且二者具有中等程度的相关性。个体通过抽样指示变量 $R \sim \text{Bernoulli}(\pi(X))$ 独立分配至不同区域。从一个大型超总体中抽取 $\{n_{\mathcal{R}}, n_{\mathcal{E}}\} = \{600, 1000\}$ 个体作为最终研究样本。治疗按 $A \sim \text{Bernoulli}(0.5)$ 分配。对于目标区域中的样本，潜在结果由 $Y(a) \sim \mathcal{N}(\beta_{a0} + X^\top \beta_{a1} + U\alpha_a, 1)$ 生成。对于辅助区域中的样本，随机比例为 $\rho = 50\%$ 的样本在 a 组中存在隐藏偏倚 b_a。偏倚样本的潜在结果由 $Y(a) \sim \mathcal{N}(-b_a + \beta_{a0} + X^\top \beta_{a1}, \epsilon)$ 生成，其中 $\epsilon = \{0.1, 0.5, 1, 1.5\}$、$b_0 = 6$ 和 $b_1 = 10$。辅助区域中其余样本不受偏倚影响，其 $Y(a) \sim \mathcal{N}(\beta_{a0} + X^\top \beta_{a1} + U\alpha_a, \epsilon)$。详细的数据生成过程总结于补充材料的附录 D.1。

我们使用基于条件化 R^2 的信噪比指标来量化信号强度，该指标定义为 $R^2_{X|U}/R^2_{U|X}$，其中 $R^2_{X|U} = \text{SSE}(X|U)/\text{SSE}(U)$ 表示在对 U 进行条件化后，由 X 解释的结局方差比例，而 $R^2_{U|X} = \text{SSE}(U|X)/\text{SSE}(X)$ 则表示在考虑 X 后，由 U 解释的剩余方差。信号比越高，表明观测协变量相对于未测量变量具有越强的预测能力。

我们从 I 类错误控制和统计功效两方面评估了 FRT 的表现。为简单起见，我们考虑了 $\epsilon = 0.5$，且隐藏偏倚仅引入对照组的情景。为了评估 I 类错误，我们在不存在区域治疗效应的零假设下生成数据，并考察隐藏偏倚幅度 $b_0 = 2$、4、6、8 逐渐增大时的经验拒绝率。为了评估检验功效，我们在一系列非零真实治疗效应下模拟数据，同时改变对照组中隐藏偏倚的程度。

在方法比较中，我们评估了六种估计量在不同借用策略和协变量调整下的表现。具体而言，我们考虑三种借用方案：NB、FB 和 CSB，采用 10 折交叉验证和绝对残差共形分数。对于每种借用方案，我们评估两个版本：一种调整观测协变量 X （仅 X），另一种同时调整 X 和 U （AllCov/IVW）。我们针对每种情景重复模拟 500 次。更多模拟结果见补充材料中的附录 D.2。6.2 模拟结果

图 2 展示了在不同精度和信号强度下的结果，其中 X以治疗效应估计量 $\hat{}\tau$的均方误差百分比（MSE%）表示，并以 NB-AllCov 估计量为标准进行标准化。我们观察到，随着 X 的信号强度或精度提高，MSE% 均会降低。这一趋势表明，当观测协变量具有信息量且结局噪声降低时，治疗效应估计会更加准确。在所有设定中，CSB 估计量的表现始终优于 NB 基线，MSE% 降幅约为 10% 至 50% 以上；在高精度、高信号的设定下，改善幅度最大。FB 估计量的表现最差，因为它们在估计中纳入了所有偏倚样本。当难以区分偏倚样本时，CSB 的表现至少不逊于 NB，这凸显了其对借用相关偏倚的稳健性。我们还观察到，与仅 X 估计量相比，AllCov/IVW 估计量的性能显著提升。这说明纳入额外协变量信息具有优势。值得注意的是，对于仅 X 估计量，即使无法获得 U，使用 X 在 X 与未测量协变量存在中等程度相关时，也能得到 MSE% 更低、更加准确的估计。这符合直觉，因为存在相关性时， X 可以部分捕捉 U中的变异。

[Figure 2 about here.]

图3 展示了随着偏倚水平不断增加而变化的经验 I 类错误率，并总结了在一系列真实治疗效应下的检验效能表现。在所有情景中，FRT 均成功将 I 类错误率控制在名义水平 0.05。检验功效

比较结果显示，三种借用策略之间存在明显且一致的优劣顺序。NB-AllCov 估计量提供了一个安全的基准，但代价是检验功效较低。FB方法表现最差，尤其是在结局异质性情况下，这是由于它不加区分地合并了存在偏倚的辅助样本。CSB估计量始终在所有设置下实现最高检验功效，因为它们仅选择性地整合结局相容数据。在所有借用方案中，AllCov/IVW估计量的统计功效均高于仅使用 X 的估计量。这种提升在 CSB-IVW估计量上尤为明显，其整体表现始终最佳。

[Figure 3 about here.]

7. 真实数据示例

7.1 数据准备

我们将所提方法应用于一项名为 POWER试验的国际多中心安慰剂对照临床试验（III期）。为作说明，我们关注的监管问题是：enobosarm 是否能有效预防和治疗晚期非小细胞肺癌患者的肌肉消耗。在北美（$R = 1$），39 名患者被随机分配接受 3 mg enobosarm（$A = 1$），30 名患者被随机分配接受安慰剂（$A = 0$）。该试验还纳入了来自不同地区的 576 名患者，包括欧洲和南美洲（$R = 0$）；这些患者的数据可作为辅助数据，用于改进北美治疗效应估计与推断。

我们将第84天瘦体重（LBM）的百分比变化作为主要终点。共有7个共享基线协变量，包括年龄、性别、ECOG状态、化疗使用、癌症分期、组织学和基线瘦体重。这些协变量反映患者特征中的关键基线特征，并作为处理协变量不可比性的基础。此外，体重下降标志是一项重要的预后因素，用于指示研究开始前6个月内体重是否下降了5%或以上。该指标在北美数据中可用，但在辅助区域的数据中缺失。

我们采用1:4 最近邻匹配程序来提高协变量的可比性。在匹配前，剔除了辅助区域中存在缺失值或协变量超出目标区域支持范围的观测值。匹配基于共享协变量进行，并保留目标区域的所有样本，以及辅助区域中的 $n_{\mathcal{E}} = 276$ 名患者；对照组和治疗组分别包括155名和121名患者。有关真实数据的更多详情，请参阅《补充材料》中的附录 E。

7.2 数据分析结果

我们采用绝对残差得分分别实现 NB、FB 和 CSB，并为每种方法使用两种协变量调整方案（仅 X 和 AllCov/IVW），以估计区域特异性平均治疗效应并开展 FRT。我们还将简单的均值差估计量（DiM）作为基准方法。此外，我们还计算了所有方法的渐近标准误、置信区间和 p 值。

图 4 展示了 CSB-IVW 下从辅助区域中选择的样本。大多数已选择的辅助样本在抽样得分和结局方面均与北美样本高度一致，这表明保形选择能够有效识别可交换辅助数据。相比之下，未选择的辅助样本往往与北美样本存在偏差，尤其是在抽样得分分布的极端位置，这说明协变量空间中的重叠有限。这一模式表明，CSB-IVW 程序成功优先选择了与目标区域更为相似的辅助数据，从而减少了潜在离群值造成的隐藏偏倚。

[图4约在此处。]

图 5 展示了不同借用策略下估计的区域特异性平均治疗效应、95%置信区间，以及渐近推断和FRT的p值。仅 X 估计量往往会高估治疗效应，这可能是因为它们未能针对协变量不匹配进行调整。相比之下，AllCov/IVW估计量通过纳入额外协变量的信息，减轻了这种膨胀。CSB-IVW估计值略低于NB-AllCov和FB-IVW估计值，这是因为它有选择地借用两个治疗组的信息，并排除了

更多结局较低的对照组患者，从而将估计值向下调整。此外，与NB和FB相比，CSB-IVW显著提高了精度，使置信区间宽度缩小约10%–45%。在渐近推断和FRT中，CSB-IVW产生的p值始终小于其不借用和完全借用的对应方法。尽管渐近p值往往小于相应的FRT p值，这反映了基于随机化的推断具有保守性，但CSB-IVW下两类p值的一致性支持所估计治疗效应的可靠性。估计的LBM增益约为2.5%，对应于具有临床意义的肌肉质量恢复改善，表明存在更强且更稳定的阳性区域特异性平均治疗效应证据。

[Figure 5 about here.]

8. 讨论

本文围绕预先指定目标区域的区域特异性平均治疗效应（RSATE），为多区域试验分析构建了形式化的估计目标，并开发了用于估计和推断的选择性信息借用框架。通过将逆方差加权估计量与保形子集选择和条件随机化检验相结合，我们的 CSB-IVW 程序在协变量偏移、协变量不匹配和结局漂移的情况下，提高了 RSATE 的效率和检验功效，同时保持有效的 I 类错误控制。这些特性使由此得到的 RSATE 推断特别适用于本地监管评估，因为该评估越来越要求明确界定估计目标，并采用透明的信息借用规则。

我们的方法框架依赖目标区域内标准的因果识别假设（一致性、随机化和正值性），并要求具备足以预测结局、从而区分可比辅助区域患者的协变量。当各地区的协变量几乎没有重叠，或结局漂移普遍存在时，所提出的

程序会自然地向仅目标估计量收缩，为监管决策提供保守但透明的借用方式。

尽管我们聚焦于多区域临床试验（MRCT）情景，但我们的方法自然延伸至增强型随机临床试验，即同时纳入外部对照和外部治疗，并能够处理协变量不匹配的外部数据，从而将Zhu 等人（2025年）的方法扩展至更广泛的实际情景，并与当前监管机构利用真实世界数据开展确证性分析的关注方向保持一致（美国食品药品监督管理局，2023年）。将 CSB 扩展至生存时间结局是可行的（Cand`等人，2023年），尤其适用于不同地区的删失情况和事件时间模式存在差异的肿瘤试验。条件随机化检验同样可以推广至这些情景（Heng 等人，2025）。关于协变量不匹配问题，重要的预后因素（如 ECOG 体能状态）在不同数据来源中的记录可能并不一致。Charlson 合并症指数等代理测量可以部分捕捉这些潜在因素。在此类情况下，当关键混杂因素无法直接观测时，近端因果方法（Tchetgen 等人，2024）为实现稳健的信息借用提供了潜在路径。

Acknowledgements

本项目由美国卫生与公众服务部（HHS）下属的美国食品药品监督管理局（FDA）通过一项为期三年、总额为 2,556,429 美元的资助项目（U01FD007934）提供支持，资金由 FDA/HHS 提供。本研究还得到由 NIH/HHS 资助的 R01AG066883 项目支持。本文内容仅代表作者观点，并不一定代表 FDA/HHS、NIH/HHS 或美国政府的官方观点，也不一定获得其认可。

Supplementary Materials

网络附录 A–E 包含技术细节、算法、证明、模拟及补充数据分析。R 代码可在此处获取：github.com/chenxi0217/RSATE-in-MRCT。

Data Availability

由于涉及参与者隐私，数据不公开。

References

Alene, M., Vansteelandt, S., and Van Lancker, K. (2025). 在考虑聚类的同时，使用协变量调整分析多中心随机试验。arXiv 预印本 arXiv:2504.12760。Alt, E. M., Chang, X., Jiang, X., Liu, Q., Mo, M., Xia, H. A., and Ibrahim, J. G. (2024). LEAP：用于从历史数据中借用信息的潜在可交换性先验。《生物统计学》80, ujae083。Angelopoulos, A. N. and Bates, S. (2023). 保形预测：简明入门。Foundations and Trends® 机器学习 16, 494–591。Barber, R. F., Cand`es, E. J., Ramdas, A., and Tibshirani, R. J. (2021). 使用刀切法+进行预测推断。The Annals of Statistics 49, 486–507。Bean, N. W., Ibrahim, J. G., and Psioda, M. A. (2021). 使用模型平均的贝叶斯多区域临床试验。Biostatistics 24, 262–276。Bean, N. W., Ibrahim, J. G., and Psioda, M. A. (2023). 具有生存时间终点的多区域临床试验的贝叶斯设计。《生物统计学》79, 3586–3598。Bean, N. W., Ibrahim, J. G., and Psioda, M. A. (2024). 多区域临床试验的贝叶斯联合模型。Biostatistics 25, 852–866。Cand`es, E., Lei, L., and Ren, Z. (2023). 保形生存分析。Journal of the Royal Statistical Society: Series B (Statistical Methodology) 85, 24–45。Chang, J. H., Russo, M., and Paul, S. (2024). 用于特征不匹配的高维回归的异质迁移学习。arXiv 预印本 arXiv:2412.18081。Chen, J., Quan, H., Binkowitz, B., Ouyang, S. P., Tanaka, Y., Li, G., Menjoge, S., and Ibia, E. (2010). 评估多区域临床试验中的治疗效应一致性：系统综述。Pharmaceutical Statistics 9, 242–253。Chen, M.-H. and Ibrahim, J. G. (2000). 回归模型的幂先验分布。

《统计科学》15, 46–60。Chen, S., Li, S., Zhang, B., 和 Ye, T.（2025）。结合实验数据与观察性数据的极小极大速率与自适应性。《因果推断杂志》13, 20240024。Chen, Y. H., Wu, Y. C., 和 Wang, M.（2009）。一种评估多区域试验区域治疗效应的贝叶斯方法。《生物制药统计学杂志》19, 900–915。Cheng, D. 和 Cai, T.（2021）。随机数据与观察性数据的自适应结合。arXiv预印本 arXiv:2111.15012。Colnet, B., Mayer, I., Chen, G., Dieng, A., Li, R., Varoquaux, G., Vert, J.-P., Josse, J., 和 Yang, S.（2024）。结合随机试验与观察性研究的因果推断方法：综述。《统计科学》39, 165–191。Dang, L. E., Tarp, J. M., Abrahamsen, T. J., Kvist, K., Buse, J. B., Petersen, M., 和 van der Laan, M.（2025）。用于混合随机对照试验与外部数据研究的实验选择器交叉验证目标最大似然估计量。《因果推断杂志》13, 20240041。Dette, H., Koletzko, L., 和 Bretz, F.（2025）。检验多区域临床试验中剂量反应的相似性。《医学统计学》44, e70255。美国食品药品监督管理局（2023）。真实世界数据：评估登记系统以支持药品和生物制品的监管决策。行业指南/技术报告，美国卫生与公众服务部，美国食品药品监督管理局。Fisher, R. A.（1935）。《实验设计》。Oliver and Boyd，爱丁堡，第1版。Gao, C. 和 Yang, S.（2023）。结合概率样本与非概率样本的预检验估计。《电子统计学杂志》17, 1492–1546。Gao, C., Yang, S., Shan, M., Ye, W., Lipkovich, I., 和 Faries, D.（2025a）。通过数据自适应借用改进随机对照试验分析。《生物计量学》112, asae069。

Gao, C., Yang, S., Shan, M., Ye, W. W., Lipkovich, I., 和 Faries, D.（2025b）。利用外部对照对随机临床试验的生存结局进行双重保护估计。载于《第42届国际机器学习会议论文集》（ICML 2025）。Guo, H., Chen, J., 和 Quan, H.（2016）。在多区域临床试验中借用相似国家的信息评估局部治疗效应。《医学统计学》35, 671–684。Guo, W., Wang, S. L., Ding, P., Wang, Y., 和 Jordan, M.（2022）。利用控制变量进行多源因果推断，以应对结局选择偏倚。《机器学习研究汇刊》。Han, L., Shen, Z., 和 Zubizarreta, J.（2023）。目标平均治疗效应的多重稳健联邦估计。《神经信息处理系统进展》36, 70453–70482。Han, P., Li, H., Park, S. K., Mukherjee, B., 和 Taylor, J. M.（2024）。整合来自异质总体的外部信息以改进线性回归模型的预测：James–Stein估计量。《生物统计学》80, ujae072。Heng, S., Zhang, J., 和 Feng, Y.（2025）。针对结局缺失的基于设计的因果推断：缺失机制、插补辅助随机化检验和协变量调整。《美国统计学会杂志》，网络先行发表。Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., 和 Sargent, D. J.（2011）。用于在临床试验中自适应纳入历史信息的分层相容先验模型与幂先验模型。《生物统计学》67, 1047–1056。Hua, K., Hong, H., 和 Wang, X.（2024）。在多区域临床试验中利用限制平均生存时间推断治疗效应及其区域修饰因子。arXiv预印本 arXiv:2404.08128。

Hung, H.、Wang, S. 和 O'Neill, R.（2010）。多区域试验设计与分析中区域差异的考量。《Pharmaceutical Statistics》9，173–178。ICH（1998）。国外临床数据可接受性中的种族因素。ICH（2017）。多区域临床试验规划与设计的一般原则。Jin, Y. 和 Cand`es, E. J.（2023）。使用保形p值进行基于预测的选择。《机器学习研究杂志》24，1–41。Kaizer, A. M.、Koopmeiners, J. S. 和 Hobbs, B. P.（2018）。基于多源可交换性的贝叶斯层级建模。Biostatistics 19，169–184。Kwiatkowski, E.、Zhu, J.、Li, X.、Pang, H.、Lieberman, G. 和 Psioda, M. A.（2024）。用于结合生存时间数据的混合对照分析的病例加权幂先验。《生物统计学》80，ujae019。Lei, J.、G'Sell, M.、Rinaldo, A.、Tibshirani, R. J. 和 Wasserman, L.（2018）。用于回归的分布无关预测推断。《美国统计学会杂志》113，1094–1111。Li, B.、Wang, Y.、Wu, A.、Ma, M.、Xiong, R. 和 Kuang, K.（2025）。在多样化环境下将随机对照试验中的因果效应推广至目标人群。载于《第42届国际机器学习会议论文集》。Li, X.、Wang, M.、Lu, F. 和 Zhou, X.-H.（2023）。利用外部对照数据提高临床试验推断的效率。《生物统计学》79，394–403。Liu, J.、Zhu, K.、Yang, S. 和 Wang, X.（2025）。二元结局混合对照试验中的稳健估计与推断：非小细胞肺癌案例研究。arXiv 预印本 arXiv:2505.00217。Liu, Y.、Zhu, K.、Han, L. 和 Yang, S.（2025）。COADVISE：随机对照试验中结合变量选择的协变量调整。《英国皇家统计学会会刊A辑：社会统计学》页码 qnaf171。

Luo, X.、Dasgupta, T.、Xie, M. 和 Liu, R. Y.（2021）。利用置信分布实施 Fisher 随机化检验：推断、组合与融合学习。《皇家统计学会会刊 B辑：统计方法学》83，777–797。Mao, G.、Yang, S. 和 Wang, X.（2025）。综合随机临床试验与真实世界数据所得右删失数据的异质性治疗效应统计推断。《生物统计学》81，ujaf131。Papadopoulos, H.、Proedrou, K.、Vovk, V. 和 Gammerman, A.（2002）。用于回归的归纳式置信机。载于《第 13 届欧洲机器学习会议》。Ren, B.、Ferrari, F.、Fortini, S.、Ventz, S. 和 Trippa, L.（2025）。利用外部数据检验随机临床试验中的试验性疗法及其生物标志物交互作用。《生物计量学》页码 asaf047。Robertson, S.、Steingrimsson, J.、Joyce, N.、Stuart, E. 和 Dahabreh, I.（2025）。多中心试验中的中心特异性因果推断——在各参与中心背景下解释试验证据。《医学研究统计方法》34，2330–2347。Romano, Y.、Patterson, E. 和 Candes, E.（2019）。保形化分位数回归。载于 Wallach, H.、Larochelle, H.、Beygelzimer, A.、d'Alch´e-Buc, F.、Fox, E. 和 Garnett, R. 编，《神经信息处理系统进展》第32卷。Rosenman, E. T.、Basse, G.、Owen, A. B. 和 Baiocchi, M.（2023）。使用收缩估计量结合观察性数据集与实验数据集。《生物计量学》79，2961–2973。Rothenh¨ausler, D.（2024）。因果参数估计的模型选择与推断。《电子统计学杂志》18，5449–5483。Schmidli, H.、Gsteiger, S.、Roychoudhury, S.、O’Hagan, A.、Spiegelhalter, D. 和 Neuen- schwander, B.（2014）。具有历史对照信息的临床试验中的稳健荟萃分析预测先验。《生物计量学》70，1023–1032。Schuler, A.、Walsh, D.、Hall, D.、Walsh, J.、Fisher, C.，代表阿尔茨海默病 C. P.，

Initiative, A. D. N., and Study, A. D. C. (2022). 通过对预后评分进行线性调整提高随机试验估计的效率。《国际生物统计学杂志》18, 329–356. Shafer, G. and Vovk, V. (2008). 保形预测教程。《机器学习研究杂志》9, 371–421. Stuart, E. A. and Rubin, D. B. (2008). 对多个对照组进行匹配并调整组间差异。《教育与行为统计学杂志》33, 279–306. Tchetgen, E. J. T., Ying, A., Cui, Y., Shi, X., and Miao, W. (2024). 近端因果推断简介。《统计科学》39, 375–390. Thall, P. F., Wathen, J. K., Bekele, B. N., Champlin, R. E., Baker, L. H., and Benjamin, R. S. (2003). 针对具有多个亚型疾病的II期试验的分层贝叶斯方法。《医学统计学》22, 763–780. Valancius, M., Pang, H., Zhu, J., Cole, S., Funk, M. J., and Kosorok, M. (2024). 利用外部对照的混合试验因果推断框架。《生物统计学》80, ujae095. van der Laan, M., Qiu, S., and van der Laan, L. (2024). 基于融入真实世界数据的随机对照试验估计平均处理效应的自适应目标最大似然估计。arXiv预印本 arXiv:2405.07186 . Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., Hobbs, B., Ibrahim, J. G., Kinnersley, N., Lindborg, S., et al. (2014). 在临床试验中使用历史对照数据评估治疗效应。Pharmaceutical Statistics 13, 41–54. Vovk, V., Gammerman, A., and Shafer, G. (2005). 《随机世界中的算法学习》，第29卷。施普林格。Wan, S.-H., Liang, H.-C., Tsou, H.-H., Wu, H.-D., and Wang, S. (2025). 关于离散随机效应模型下多区域临床试验总体治疗效应的估计

《医学研究统计方法》34, 735–750. Williams, N., Rudolph, K., and D´ıaz, I. (2025). 使用融合随机试验和缺失的效应修饰因子对最优治疗规则进行非参数估计。arXiv预印本 arXiv:2506.10863 . Wu, L. and Yang, S. (2022). 结合实验性研究和观察性研究的异质性治疗效应整合 R-学习器。载于首届因果学习与推理会议论文集，《机器学习研究论文集》第177卷，第904–926页。Xu, Q. and Qu, A. (2025). 用于异质数据整合的表征检索学习。arXiv预印本 arXiv:2503.09494 . Yang, P., Zhao, Y., Nie, L., Vallejo, J., and Yuan, Y. (2023). SAM：在临床试验中动态借用历史数据的信息的自适应混合先验。《生物统计学》79, 2857–2868. Yang, S. and Ding, P. (2020). 结合多个观察性数据来源估计因果效应。《美国统计学会杂志》115, 1540–1554. Yang, S., Gao, C., Zeng, D., and Wang, X. (2023). 随机试验与真实世界数据的弹性整合分析，用于治疗异质性估计。《皇家统计学会会刊 B辑：统计方法学》85, 575–596. Yang, S., Liu, S., Zeng, D., and Wang, X. (2024). 治疗效应异质性和混杂函数的数据融合方法。《伯努利分布》即将刊出。Ye, X., Yang, S., Wang, X., and Liu, Y. (2025). 对存在删失和隐性混杂的高维随机对照试验与真实世界数据进行整合分析。《寿命数据分析》31, 473–497. Zeng, Z., Kennedy, E. H., Bodnar, L. M., and Naimi, A. I. (2025). 高效泛化

以及运输。《统计科学》40，495–514。Zhang, Y. 和 Zhao, Q.（2023）。什么是随机化检验？《美国统计学会杂志》118，2928–2942。Zheng, L. 和 Zelen, M.（2008）。多中心临床试验：随机化与辅助统计量。《应用统计年鉴》2，582–600。Zhu, K. 和 Liu, H.（2023）。配对交换再随机化。《生物统计学》79，2127–2142。Zhu, K. 和 Liu, H.（2024）。对 Jacob Fiksel 所著“关于基于精确随机化的协变量调整置信区间”的答辩。《生物统计学》80，ujae052。Zhu, K.、Yang, S. 和 Wang, X.（2025）。提升混合对照试验的统计有效性和检验功效：一种采用保形选择性借用的随机化推断方法。载于《第42届国际机器学习会议论文集》。Zhuang, H.、Wang, X. 和 George, S. L.（2024）。评估多区域随机临床试验中的治疗效应异质性。《生物制药研究统计学》17，315–322。Zou, H.（2006）。自适应套索及其预言机性质。《美国统计学会杂志》101，1418–1429。

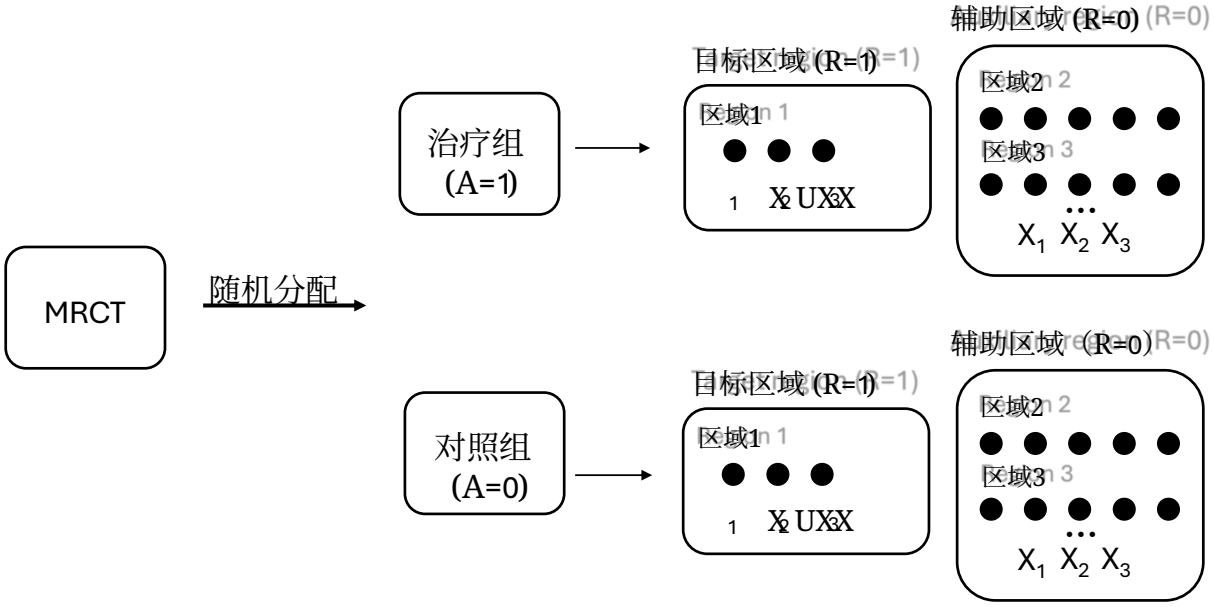

图1. 包含目标区域和辅助区域的多区域随机临床试验的数据结构。

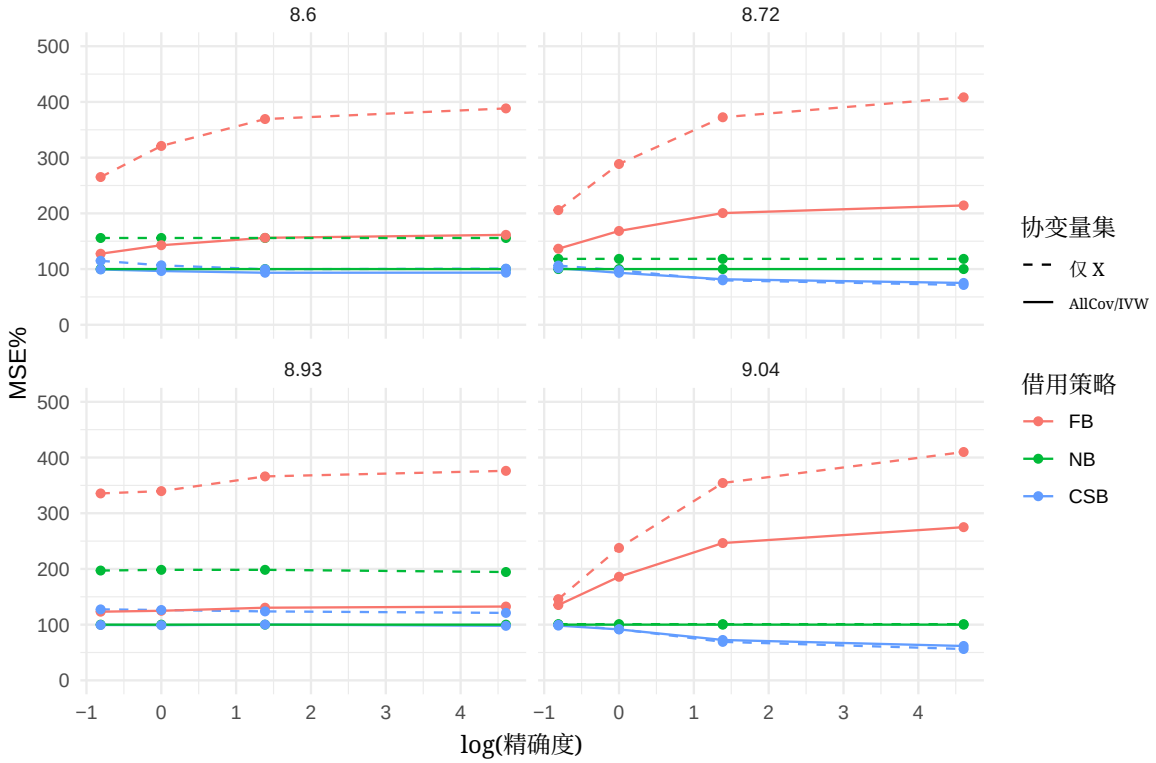

图 2。 X 与相关的 X 和 U的不同信号下，均方误差百分比与 log(精确度) 的关系。

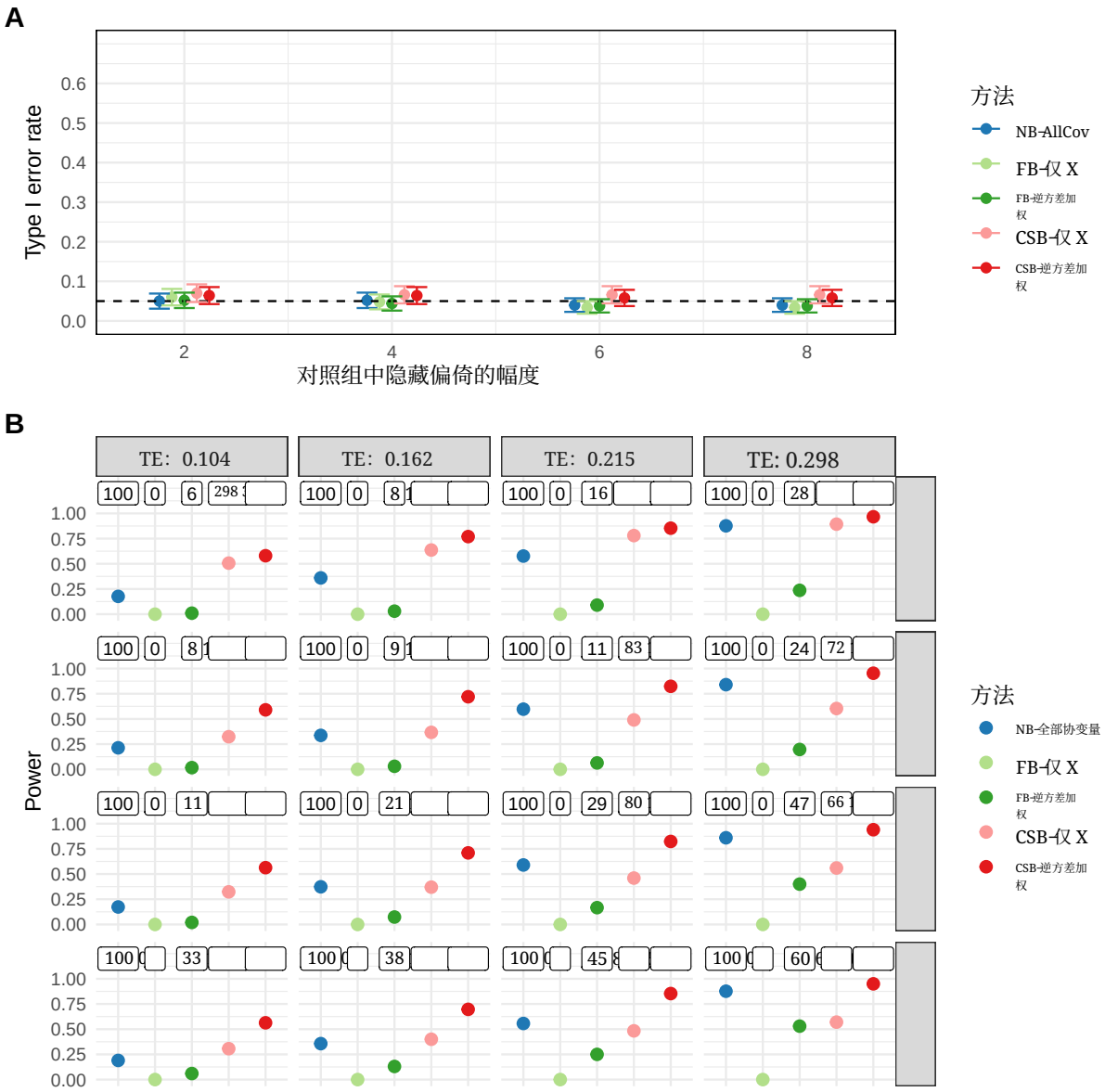

图 3. 在不同隐藏偏倚程度下，FRT 的第一类错误率；在不同真实治疗效应和隐藏偏倚下的统计功效。

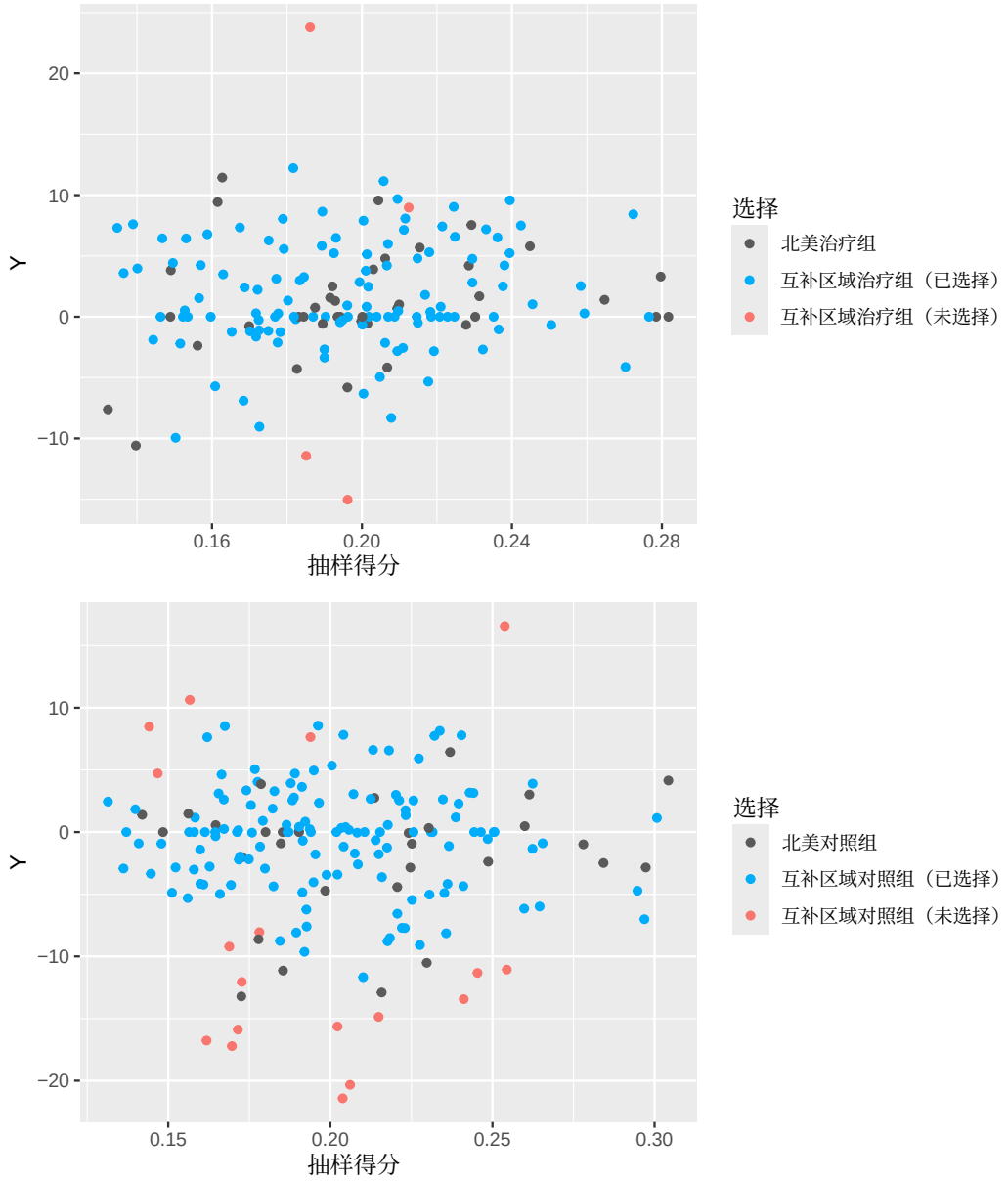

图4. 结局（第84天的瘦体重）与抽样得分分布

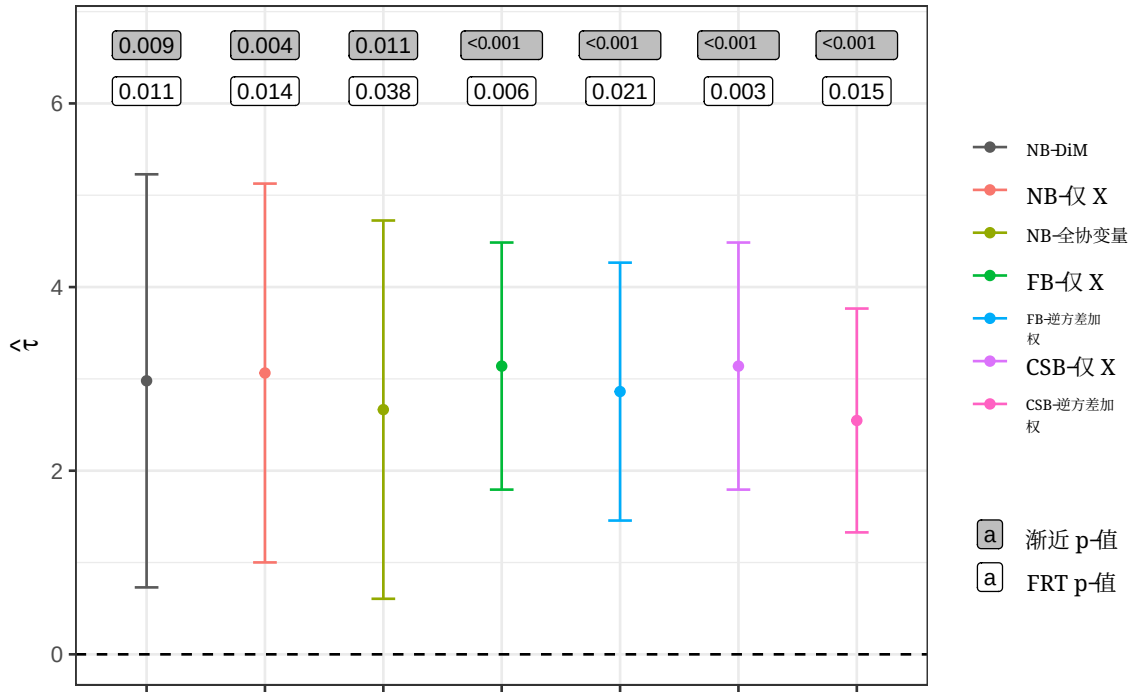

图 5。不同借用策略下估计的治疗效应和 p 值的比较。上方（下方）数字分别报告渐近（FRT）p 值。

表 1*Observed data structure. "✓" and "?" indicate observed and unobserved, respectively*

	ID	处理	协变量	潜在结果			观测结果
		A	X	U	$Y(1)$	$Y(0)$	Y
目标	1	1	✓	✓	✓	?	✓
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$n_{\mathcal{R}}$	0	✓	✓	?	✓	✓
辅助	$n_{\mathcal{R}}+1$	1	✓	?	✓	?	✓
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	n	0	✓	?	?	✓	✓

《多区域临床试验中区域特异性治疗效应推断的选择性信息借用》的补充材料

Chenxi Li[1]，Ke Zhu[1,2],Shu Yang[2] 和 Xiaofei Wang[1,*]

[1]杜克大学生物统计学与生物信息学系，美国北卡罗来纳州达勒姆 27710。[2]北卡罗来纳州立大学统计学系，美国北卡罗来纳州罗利 27695。*电子邮箱：xiaofei.wang@duke.edu

附录 A 给出了正文中介绍的估计方法所对应的渐近方差估计量和推断程序。附录 B 概述了自适应选择阈值所用的算法。附录 C 证明了定理1，即 Fisher 随机化检验的有效性。附录 D 总结了模拟研究中使用的数据生成机制和参数，并给出了更多模拟结果。附录 E 报告了真实数据应用的补充结果。附录 F 将 FB-IVW 扩展到具有不同共享协变量的多个辅助区域：通过区域特异性识别并对各区域的估计量进行最优组合，推断目标区域的治疗效应。

附录 A：半参数有效估计量与渐近推断

令$z_{1-\alpha/2}$ 表示标准正态分布的（$1-\alpha/2$）分位数，且 $\pi_{\mathcal{R}}$ $=\mathbb{E}\{\pi(X)\}$ 表示目标区域数据的样本比例。令 $O=(R,A,Y,X,U)$表示观测数据，其中 U 对 $R=0$部分缺失。

A.1 仅目标区域分析（NB-AllCov）

Let

$$\xi_a(O)=\mu_a(X,U)+\frac{\mathbb{I}(R=1,A=a)}{e_{a|1}(X)}\{Y-\mu_a(X,U)\}$$

and $\hat{\xi}(O_i)$ 表示带有已估计干扰函数的 $\xi_a(O)$的经验版本，即 s,

$$\hat{\xi}_a(O_i)=\hat{\mu}_a^{\mathrm{NB}}(X_i,U_i)+\frac{\mathbb{I}(R_i=1,A_i=a)}{e_{a|1}(X_i)}\{Y_i-\hat{\mu}_a^{\mathrm{NB}}(X_i,U_i)\}.$$

$\sqrt{n}\hat{\tau}^{\text{NB-AllCov}}$ 的方差估计量为

$$\begin{aligned}\hat{V}^{\text{NB-AllCov}}&=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{R_i}{\pi_{\mathcal{R}}}\left\{\hat{\xi}_1(O_i)-\hat{\xi}_0(O_i)-\hat{\tau}^{\text{NB-AllCov}}\right\}\right]^2\\&=\frac{1}{n{\pi_{\mathcal{R}}}^2}\sum_{i:R_i=1}\left\{\hat{\xi}_1(O_i)-\hat{\xi}_0(O_i)-\hat{\tau}^{\text{NB-AllCov}}\right\}^2.\end{aligned}$$

渐近置信区间为

$$\left[\hat{\tau}^{\text{NB-AllCov}} - z_{1-\alpha/2}\sqrt{\hat{V}^{\text{NB-AllCov}}/n},\ \hat{\tau}^{\text{NB-AllCov}} + z_{1-\alpha/2}\sqrt{\hat{V}^{\text{NB-AllCov}}/n}\right].$$

A.2 利用共享协变量从辅助区域借用信息（FB-Xonly）

Let

$$\phi_a(O) = \frac{R}{\pi_{\mathcal{R}}}\mu_a(X) + \frac{\pi(X)}{\pi_{\mathcal{R}}}\frac{\mathbb{I}(A=a)}{e_a(X)}\{Y - \mu_a(X)\},$$

ˆ和 $\phi_a(O_i)$ 表示使用估计的干扰函数得到的 $\phi_a(O)$ 的经验版本，即

$$\hat{\phi}_a(O_i) = \frac{R_i}{\pi_{\mathcal{R}}}\hat{\mu}_a^{\text{FB}}(X_i) + \frac{\hat{\pi}^{\text{FB}}(X_i)}{\pi_{\mathcal{R}}}\frac{\mathbb{I}(A_i=a)}{e_a(X_i)}\{Y_i - \hat{\mu}_a^{\text{FB}}(X_i)\}.$$

$\sqrt{n}\hat{\tau}^{\text{FB-Xonly}}$ 的方差估计量为

$$\hat{V}^{\text{FB-Xonly}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\phi}_1(O_i) - \hat{\phi}_0(O_i) - \frac{R_i}{\pi_{\mathcal{R}}}\hat{\tau}^{\text{FB-Xonly}}\right\}^2.$$

渐近置信区间为

$$\left[\hat{\tau}^{\text{FB-Xonly}} - z_{1-\alpha/2}\sqrt{\hat{V}^{\text{FB-Xonly}}/n},\ \hat{\tau}^{\text{FB-Xonly}} + z_{1-\alpha/2}\sqrt{\hat{V}^{\text{FB-Xonly}}/n}\right].$$

A.3 使用逆方差加权从辅助区域借用信息（FB-IVW）

Let

$$\psi_a(O) = \frac{R}{\pi_{\mathcal{R}}}\mu_a + \frac{\pi(X)}{\pi_{\mathcal{R}}}\frac{\mathbb{I}(A=a)}{e_a(X)}\{Y - \mu_a\},$$

and $\hat{\psi}_a(O_i)$ d 记估计了干扰函数的 $\psi_a(O)$的经验版本为，即:

$$\hat{\psi}_a(O_i) = \frac{R_i}{\pi_{\mathcal{R}}}\hat{Y}_i^{\text{FB}} + \frac{\hat{\pi}^{\text{FB}}(X_i)}{\pi_{\mathcal{R}}}\frac{\mathbb{I}(A_i=a)}{e_a(X_i)}\{Y_i - \hat{Y}_i^{\text{FB}}\}.$$

其中

$$\hat{Y}_i^{\text{FB}} = \begin{cases} \dfrac{v_{\text{FB}}}{v_{\text{NB}}+v_{\text{FB}}}\hat{\mu}_a^{\text{NB}}(X_i,U_i) + \dfrac{v_{\text{NB}}}{v_{\text{NB}}+v_{\text{FB}}}\hat{\mu}_a^{\text{FB}}(X_i), & R_i = 1, \\ \hat{\mu}_a^{\text{FB}}(X_i), & R_i = 0. \end{cases}$$

对于$\sqrt{n}\hat{\tau}^{\text{FB-IVW}}$ 的方差估计量为

$$\hat{V}^{\text{FB-IVW}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\psi}_1(O_i) - \hat{\psi}_0(O_i) - \frac{R_i}{\pi_R}\hat{\tau}^{\text{FB-IVW}}\right\}^2.$$

渐近置信区间为

$$\left[\hat{\tau}^{\text{FB-IVW}} - z_{1-\alpha/2}\sqrt{\hat{V}^{\text{FB-IVW}}/n},\ \hat{\tau}^{\text{FB-IVW}} + z_{1-\alpha/2}\sqrt{\hat{V}^{\text{FB-IVW}}/n}\right].$$

REMARK 1: 注意，通过重新整理各项，我们可以得到等价表示

$$\begin{aligned}\hat{\theta}_a^{\text{FB-IVW}} = &\frac{1}{n_{\mathcal{R}}}\sum_{i=1}^{n} R_i\,\hat{\pi}^{\text{FB}}(X_i)\,\frac{\mathbb{I}(A_i=a)}{e_a(X_i)}\,Y_i \\ &+ \frac{1}{n_{\mathcal{R}}}\sum_{i=1}^{n} R_i\left\{1-\hat{\pi}^{\text{FB}}(X_i)\frac{\mathbb{I}(A_i=a)}{e_a(X_i)}\right\}\hat{Y}_i^{\text{FB}} \\ &+ \frac{1}{n_{\mathcal{R}}}\sum_{i=1}^{n}(1-R_i)\,\hat{\pi}^{\text{FB}}(X_i)\frac{\mathbb{I}(A_i=a)}{e_a(X_i)}\left(Y_i - \hat{Y}_i^{\text{FB}}\right).\end{aligned}$$

第一项对应于基于目标区域得到的估计量。在假设2下，$\hat{Y}_i^{\text{FB}}$能够一致估计条件均值结局，这意味着其余各项可视为渐近均值为零的增广项。从这个意义上说，$\hat{\theta}_a^{\text{FB-IVW}}$可被视为仅目标估计量加上一个均值为零的增广项；这一概念与Yang 和 Ding (2020)提出的效率增广思想相关，即通过加入一个零估计量来提高仅目标估计量的效率。

附录 B: γ_a 的自适应选择阈值算法

Algorithm 1 Adaptive Selection Threshold of γ_a

Input: Threshold grid $\mathcal{G} = \{0, 0.1, \ldots, 1\}$, number of bootstrap samples $L = 100$.

for $\gamma_a \in \mathcal{G}$ **do**
  Compute $\hat{\theta}_a^{\text{CSB-IVW}}(\gamma_a)$ from the original sample.
  for $l = 1, \ldots, L$ **do**
    Compute $\hat{\theta}_a^{\text{CSB-IVW}(l)}(\gamma_a)$ from the l-th bootstrap sample.
  end for
end for
for $\gamma_a \in \mathcal{G} \setminus \{1\}$ **do**
  $\widehat{\mathbb{V}}\{\hat{\theta}_a^{\text{CSB-IVW}}(\gamma_a) - \hat{\theta}_a^{\text{NB-AllCov}}\} = (L-1)^{-1} \sum_{l=1}^{L} \Big[\{\hat{\theta}_a^{\text{CSB-IVW}(l)}(\gamma_a) - \hat{\theta}_a^{\text{NB-AllCov}(l)}\}$
  $- L^{-1} \sum_{l'=1}^{L} \{\hat{\theta}_a^{\text{CSB-IVW}(l')}(\gamma_a) - \hat{\theta}_a^{\text{NB-AllCov}(l')}\} \Big]^2$
  $\widehat{\mathbb{V}}\{\hat{\theta}_a^{\text{CSB-IVW}}(\gamma_a)\} = (L-1)^{-1} \sum_{l=1}^{L} \Big\{ \hat{\theta}_a^{\text{CSB-IVW}(l)}(\gamma_a) - L^{-1} \sum_{l'=1}^{L} \hat{\theta}_a^{\text{CSB-IVW}(l')}(\gamma_a) \Big\}^2$
  $\widehat{\text{MSE}}_a(\gamma_a) = \{\hat{\theta}_a^{\text{CSB-IVW}}(\gamma_a) - \hat{\theta}_a^{\text{NB-AllCov}}\}^2 - \widehat{\mathbb{V}}\{\hat{\theta}_a^{\text{CSB-IVW}}(\gamma_a) - \hat{\theta}_a^{\text{NB-AllCov}}\} + \widehat{\mathbb{V}}\{\hat{\theta}_a^{\text{CSB-IVW}}(\gamma_a)\}$
end for
$\widehat{\text{MSE}}_a(1) = (L-1)^{-1} \sum_{l=1}^{L} \Big\{ \hat{\theta}_a^{\text{NB-AllCov}(l)} - L^{-1} \sum_{l'=1}^{L} \hat{\theta}_a^{\text{NB-AllCov}(l')} \Big\}^2$
$\gamma_a^* = \arg\min_{\gamma_a \in \mathcal{G}} \widehat{\text{MSE}}_a(\gamma_a)$
Output: γ_a^*

附录 C：定理1的证明

在 H_0 下，插补潜在结果与真实潜在结果相同。因此，给定与辅助区域观测 $\boldsymbol{A}_{\mathcal{E}}$ 相对应的子分配向量，$T^* \equiv T(\boldsymbol{A}^*)$ 的分布与 $T \equiv T(\boldsymbol{A})$ 的分布相同；在构建目标区域内处理分配 $\boldsymbol{A}$ 的随机化分布时，该子分配向量保持固定。采用简化记号，我们有

$$\mathbb{P}_{\boldsymbol{A}}\big(p^{\text{FRT}} \leqslant \alpha \big| \boldsymbol{A}_{\mathcal{E}}\big) = \mathbb{P}_{\boldsymbol{A}}\big\{\mathbb{P}_{\boldsymbol{A}^*}(T^* \geqslant T) \leqslant \alpha \big| \boldsymbol{A}_{\mathcal{E}}\big\}.$$

在有限样本中，$\boldsymbol{A}$ 只能取有限个值，这意味着 T 也只能取有限个值。假设这些值为

$$T_1 > \cdots > T_m > \cdots > T_M,$$

and

$$\mathbb{P}_{\boldsymbol{A}}(T = T_m|\boldsymbol{A}_{\mathcal{E}}) = \mathbb{P}_{\boldsymbol{A}^*}(T^* = T_m|\boldsymbol{A}_{\mathcal{E}}) = \alpha_m, \qquad m = 1, \ldots, M.$$

对于 $T \in \{T_1, \ldots, T_M\}$，我们有 $\alpha_1 \leqslant \mathbb{P}_{\boldsymbol{A}^*}(T^* \geqslant T) \leqslant \sum_{m=1}^{M} \alpha_m = 1$。如果 $0 < \alpha < \alpha_1$，则有

$$\mathbb{P}_{\boldsymbol{A}}\big(p^{\mathrm{FRT}} \leqslant \alpha\big|\big) = \mathbb{P}_{\boldsymbol{A}}\big\{\mathbb{P}_{\boldsymbol{A}^*}(T^* \geqslant T) \leqslant \alpha\big|\boldsymbol{A}_{\mathcal{E}}\big\} = 0 \leqslant \alpha.$$

若 $\alpha_1 \leqslant \alpha < 1$，则存在 $\exists\, \tilde{M} \in \{1, \ldots, M-1\}$ 使得 $\sum_{m=1}^{M} \alpha_m \leqslant \alpha$ 且 $\sum_{m=1}^{M+1} \alpha_m > \alpha$。于是，我们有

$$\mathbb{P}_{\boldsymbol{A}}\big(p^{\mathrm{FRT}} \leqslant \alpha\big|\boldsymbol{A}_{\mathcal{E}}\big) = \mathbb{P}_{\boldsymbol{A}}\big\{\mathbb{P}_{A^*}(T^* \geqslant T) \leqslant \alpha\big|\boldsymbol{A}_{\mathcal{E}}\big\} = \mathbb{P}_{\boldsymbol{A}}\{T \in \{T_1, \ldots, T_{\tilde{M}}\}|\boldsymbol{A}_{\mathcal{E}}\} = \sum_{m=1}^{\tilde{M}} \alpha_m \leqslant \alpha.$$

I若不同的 $\boldsymbol{A} \in \mathcal{A}$ 对应的 f $T(\boldsymbol{A})$ 取值互不相同，则 p^{FRT} 服从均匀分布 ：

$$\mathbb{P}_{\boldsymbol{A}}\left(p^{\mathrm{FRT}} = \frac{a}{|\mathcal{A}|}|\boldsymbol{A}_{\mathcal{E}}\right) = \frac{1}{|\mathcal{A}|}, \qquad a = 1, \ldots, |\mathcal{A}|.$$

因此，我们有

$$\mathbb{P}_{\boldsymbol{A}}\big(p^{\mathrm{FRT}} \leqslant \alpha\big|\boldsymbol{A}_{\mathcal{E}}\big) = \frac{\lfloor \alpha|\mathcal{A}| \rfloor}{|\mathcal{A}|} > \frac{\alpha|\mathcal{A}| - 1}{|\mathcal{A}|} = \alpha - \frac{1}{|\mathcal{A}|}.$$

我们可以移除对 $\boldsymbol{A}_{\mathcal{E}}$ 的条件化，方法是对其随机化分布取平均。

根据全概率定律，我们有

$$\mathbb{P}_{\boldsymbol{A}}\big(p^{\mathrm{FRT}} \leqslant \alpha\big) = \sum_{\boldsymbol{A}_{\mathcal{E}}} \mathbb{P}_{\boldsymbol{A}}\big(p^{\mathrm{FRT}} \leqslant \alpha, \boldsymbol{A}_{\mathcal{E}}\big) = \sum_{\boldsymbol{A}_{\mathcal{E}}} \mathbb{P}_{\boldsymbol{A}}\big(p^{\mathrm{FRT}} \leqslant \alpha \mid \boldsymbol{A}_{\mathcal{E}}\big)\, \mathbb{P}(\boldsymbol{A}_{\mathcal{E}}),$$

其中求和遍历所有可能的随机辅助数据。

由（i）部分可知，在尖锐零假设 H_0 下，对于每个 $\boldsymbol{A}_{\mathcal{E}}$，

$$\mathbb{P}_{\boldsymbol{A}}\big(p^{\mathrm{FRT}} \leqslant \alpha \mid \boldsymbol{A}_{\mathcal{E}}\big) \leqslant \alpha.$$

将此界代入前面的公式可得

$$\mathbb{P}_{\boldsymbol{A}}\big(p^{\mathrm{FRT}} \leqslant \alpha\big) = \sum_{\boldsymbol{A}_{\mathcal{E}}} \mathbb{P}_{\boldsymbol{A}}\big(p^{\mathrm{FRT}} \leqslant \alpha \mid \boldsymbol{A}_{\mathcal{E}}\big)\, \mathbb{P}(\boldsymbol{A}_{\mathcal{E}}) \leqslant \sum_{\boldsymbol{A}_{\mathcal{E}}} \alpha\, \mathbb{P}(\boldsymbol{A}_{\mathcal{E}}) = \alpha \sum_{\boldsymbol{A}_{\mathcal{E}}} \mathbb{P}(\boldsymbol{A}_{\mathcal{E}}) = \alpha.$$

附录 D：更多模拟细节

D.1：模拟数据生成参数

表S1 总结了数据生成过程。

[表 1 见此处。]

D.2：不同信号下均方误差百分比与 log(精确度) 的模拟结果，其中 X 具有独立的 X 和 U。

图S1 展示了六种估计量在观测协变量的精度和信号强度不同的情况下的表现 X 其中 X 和 U 相互独立。

[Figure 1 about here.]

附录 E：关于真实数据的更多细节

E.1：各地区按治疗组划分的患者基线特征

匹配后各研究臂的详细基线特征汇总于表S2。

[Table 2 about here.]

E.2：协变量平衡

图S2 显示，匹配后协变量和抽样得分的分布平衡显著改善。然而，年龄和基线瘦体重等变量仍存在不平衡，北美患者通常年龄较大，且基线瘦体重较低。若不舍弃北美样本，这些残余不平衡无法通过匹配解决，因此有必要采用第 3 节和第 4 节提出的双重稳健估计量。

[图 2 置于此处。]

F：扩展至具有不同协变量集的多个辅助区域

F.1 数据结构

令区域指示变量满足 $R \in \{1\} \cup \mathcal{B}_0$，其中 $R = 1$ 表示目标区域，$\mathcal{B}_0 = \{r_1, \ldots, r_S\}$ 表示 S 辅助区域。观测数据仍为

$$\{R_i, X_i, U_i, A_i, Y_i\}_{i=1}^n,$$

其中，仅当 $R_i = 1$ 时才观测到 U_i 。对于每个辅助区域 $r \in \mathcal{B}_0$，令 $X^{(r)} \subseteq X$ 表示在目标区域（$R = 1$）和辅助区域 r中均有观测且定义良好的基线协变量子集。这些共享协变量集可以因辅助区域而异。正文中的协变量向量 X 可视为至少在一个辅助区域中观测到的所有基线协变量的集合，而从区域 r 借用信息时，仅以 $X^{(r)}$为条件。我们关注的估计目标仍为目标区域的区域特异性平均处理效应 $\tau = \mathbb{E}\{Y(1) - Y(0) \mid R = 1\}$。

F.2 区域特异性识别与 FB-IVW 估计

我们将假设2推广为区域特异性形式。

Assumption 3 （区域特异性条件均值可交换性）：对于每个辅助区域 $r \in \mathcal{B}_0$ 和治疗组 $a \in \{0, 1\}$,

$$\mathbb{E}\{Y(a) \mid R = r, X^{(r)}\} = \mathbb{E}\{Y(a) \mid R = 1, X^{(r)}\}.$$

该假设仅要求目标区域与某个给定辅助区域之间在给定其共享协变量 $X^{(r)}$ 的条件下满足条件均值可交换性，而不要求所有区域都具有同一个协变量集。

在假设 3下，使用目标区域和辅助区域 r 的数据，我们仅使用目标区域数据拟合$\widehat{\mu}_a^{\mathrm{NB}}(X, U)$，并使用来自地区$R = 1$ 和 $R = r$的合并数据拟合 $\widehat{\mu}_{a}^{\mathrm{FB}(r)}(X^{(r)})$。

逆方差加权预测定义如下：

$$\hat{Y}_i^{\mathrm{FB}(r)} = \begin{cases} \dfrac{v_{\mathrm{FB}}^{(r)}}{v_{\mathrm{NB}} + v_{\mathrm{FB}}^{(r)}} \hat{\mu}_a^{\mathrm{NB}}(X_i, U_i) + \dfrac{v_{\mathrm{NB}}}{v_{\mathrm{NB}} + v_{\mathrm{FB}}^{(r)}} \hat{\mu}_a^{\mathrm{FB}^{(r)}}(X_i^{(r)}), & R_i = 1, \\ \hat{\mu}_a^{\mathrm{FB}^{(r)}}(X_i^{(r)}), & R_i = r, \end{cases}$$

其中

$$v_{\mathrm{NB}} = \frac{\sum_{i=1}^n \mathbb{I}(A_i = a, R_i = 1)\{Y_i - \hat{\mu}_a^{\mathrm{NB}}(X_i, U_i)\}^2}{\sum_{i=1}^n \mathbb{I}(A_i = a, R_i = 1)},$$

$$v_{\mathrm{FB}}^{(r)} = \frac{\sum_{i=1}^n \mathbb{I}(A_i = a, R_i \in \{1, r\})\{Y_i - \hat{\mu}_a^{\mathrm{FB}(r)}(X_i^{(r)})\}^2}{\sum_{i=1}^n \mathbb{I}(A_i = a, R_i \in \{1, r\})}.$$

与第3节中的估计量类似，相应的区域特异性全借用逆方差加权估计量为

$$\hat{\theta}_a^{\text{FB-IVW}(r)} = \frac{1}{n_R} \sum_{i=1}^n \left[\mathbb{I}(R_i = 1)\hat{Y}_i^{\mathrm{FB}(r)} + \hat{\pi}^{FB(r)}(X_i^{(r)}) \frac{\mathbb{I}(A_i = a,\ R_i \in \{1, r\})}{e_a(X_i^{(r)})} \{Y_i - \hat{Y}_i^{\mathrm{FB}(r)}\} \right].$$

当所有辅助区域被合并并共享同一协变量集时，$\hat{\theta}_a^{\text{FB-IVW}(r)}$退化为第3节中定义的估计量$\hat{\theta}_a^{\text{FB-IVW}}$。在有多个辅助区域可用时，每个区域 $r \in \mathcal{B}_0$ 都会产生一个区域特异性估计量$\hat{\theta}_a^{\text{FB-IVW}(r)}$。然后，使用来自区域 r 的数据得到的相应区域特异性平均治疗效应为$\hat{\tau}^{\text{FB-IVW}(r)} = \hat{\theta}_1^{\text{FB-IVW}(r)} - \hat{\theta}_0^{\text{FB-IVW}(r)}$。

我们还可以使用逆方差加权（Gao 等，2025）进一步组合这些估计量。令 $\hat{\tau}$FB-IVW($\mathrm{r_1}$:$\mathrm{r_S}$) $= \left(\hat{\tau}^{\text{FB-IVW}(r_1)} \ldots, \hat{\tau}\text{FB-IVW}(r_S)\right)^\top$ 表示与辅助区域 $B_0 = \{r_1, \ldots, r_S\}$ 对应的按区域拼接的 FB-IVW 估计量，则有 $n^{1/2}\left(\hat{\tau}\text{FB-IVW}(\mathrm{r_1}:\mathrm{r_S}) - \tau\right) \to N(0, \Sigma)$，其中 $\Sigma = \mathbb{E}[\psi^{(r_1:r_S)}(O)\{\psi^{(r_1:r_S)}(O)\}^\top]$，且 $\psi^{(r_1:r_S)}$ 表示 $\hat{\tau}^{\text{FB-IVW}(\mathrm{r_1}:\mathrm{r_S})}$ 的影响函数。因此，整合所有辅助区域信息的最终整合估计量为 $\hat{\tau}\text{FB-IVW}(final) = \hat{d}^\top \hat{\tau}\text{FB-IVW}(\mathrm{r_1}:\mathrm{r_S}) \to N\left(0, \hat{d}^\top \Sigma \hat{d}\right)$，其中，用于组合 $\hat{\tau}\text{FB-IVW}(\mathrm{r_1}:\mathrm{r_S})$ 的最优整合权重 $\hat{d}$ 可通过最小化整合后方差获得：min $d^\top \Sigma d$，满足 $d^\top 1_S = 1$。通过拉格朗日乘数法，我们可以证明 $\hat{d} = \left\{1_S^\top \Sigma^{-1} 1_S\right\}^{-1} \Sigma^{-1} 1_S$ 是最优组合权重。

F.3 按区域进行保形选择性借用

当证据表明假设3 不成立时，我们可以按照第4.1节对每个辅助区域应用CSB。对于辅助观测 j 且 $R_j = r \in \mathcal{B}_0$ 和 $A_j = a$的情况，通过将保形 p-值 p_j 与目标区域观测$X_j^{(r)}$ ， Y_j进行比较$\{(X_i,\ Y_i)\colon\ R_i = 1,\ A_i = a\}$，即可得到该值。对于每个治疗组 a和辅助区域 r，定义选定辅助集

$$\mathcal{E}_a^{(r)}(\gamma_a^{(r)}) = \{j : R_j = r,\ A_j = a,\ p_j \geqslant \gamma_a^{(r)}\}.$$

随后，使用目标区域数据和合并选定辅助观测 $\bigcup_{r\in\mathcal{B}_0} \mathcal{E}_a^{(r)}$（$\gamma_a^{(r)}$）构建CSB–IVW估计量，并在 $X^{(r)}$上按区域定义结局回归、抽样得分和IVW预测。可使用与正文相同的MSE引导程序，分别为各辅助区域选择阈值 $\gamma_a^{(r)}$。

参考文献

Gao, C., Yang, S., Shan, M., Ye, W., Lipkovich, I., 和 Faries, D. (2025)。通过数据自适应借用改进随机对照试验分析。《生物计量学》112，asae069。

Yang, S. 和 Ding, P.（2020）。结合多个观测数据源估计因果效应。《美国统计学会杂志》115，1540–1554。

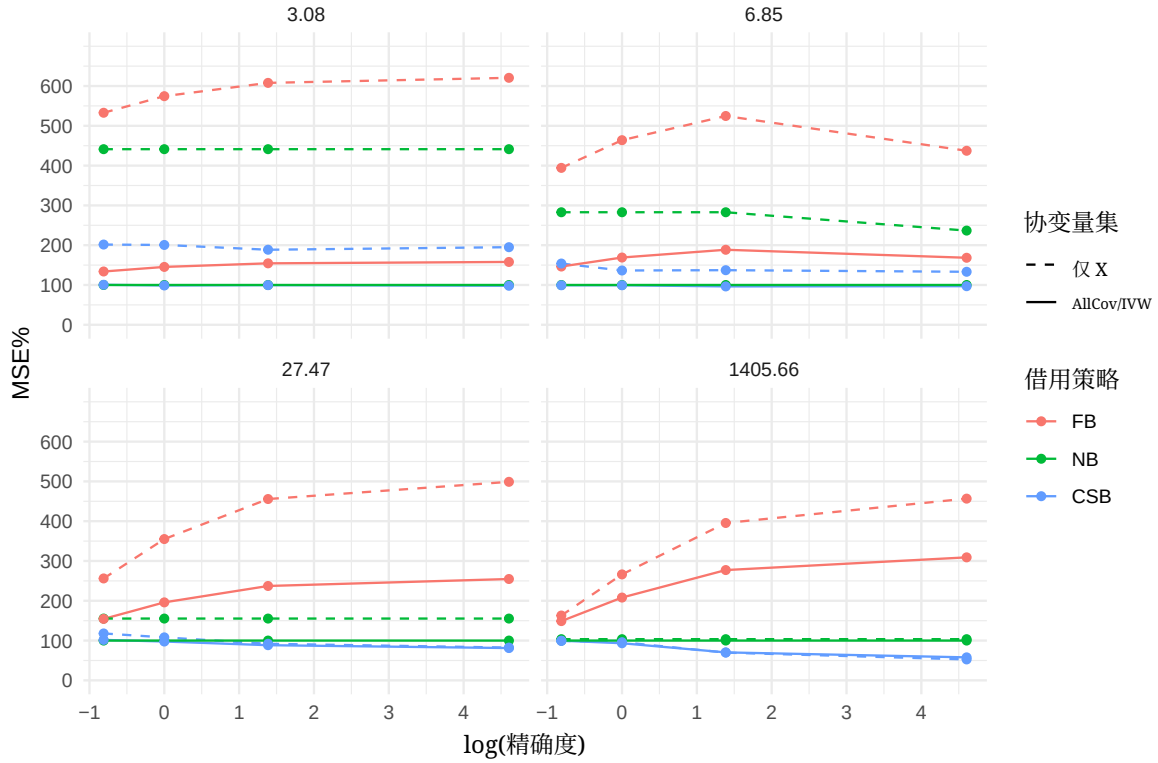

图 S1. 在不同信号下， X 具有独立的 X 和U时，均方误差百分比与log(精确度)的关系。

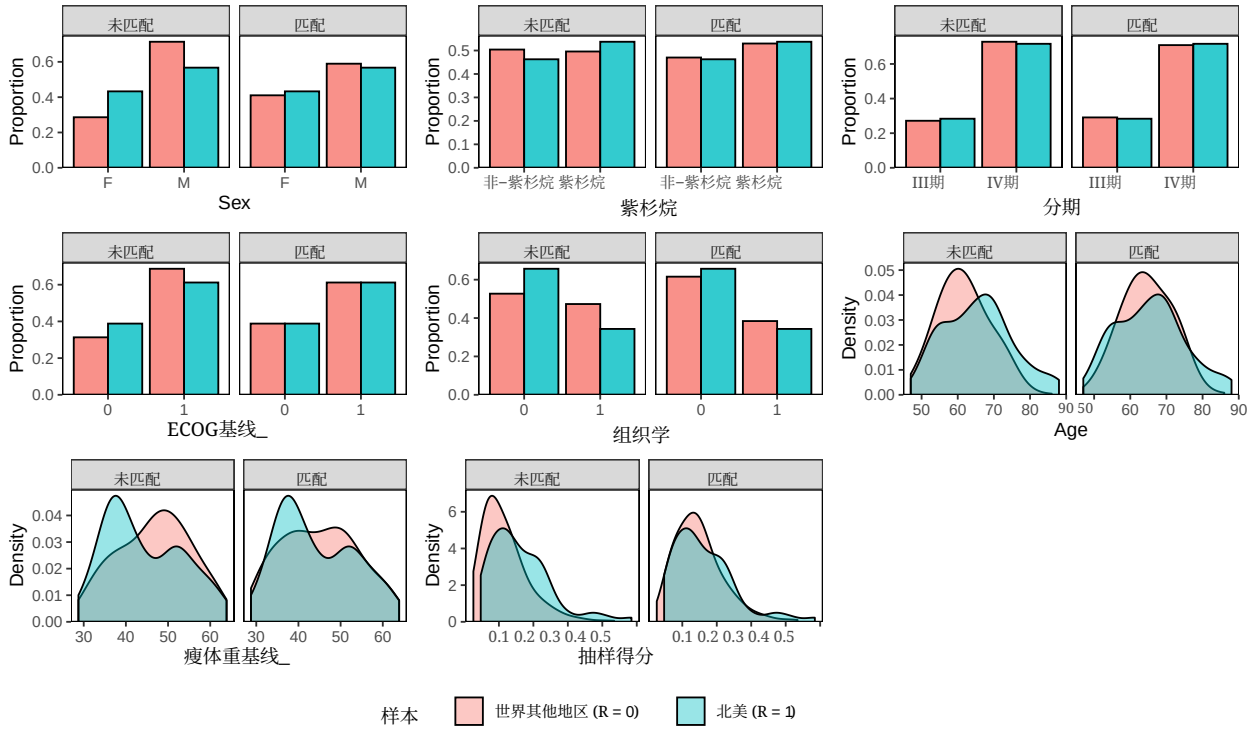

图 S2. 北美 ($R = 1$) 与辅助地区 ($R = 0$) 在基线协变量和估计抽样得分 $\mathbb{P}(R = 1|X)$ 方面的分布平衡（未匹配和匹配）。

表 S1

Simulation data-generating parameters.

参数	描述	数值 / 规格
$\{n_{\mathcal{R}}, n_{\mathcal{E}},\}$	目标区域和辅助区域的样本量	$\{600,\ 1000\}$
$X, U = (X_1, X_2, U)$	协变量	(i) $\mathcal{MVN}(\left(\begin{smallmatrix}0\\0\\2\end{smallmatrix}\right), \left(\begin{smallmatrix}1 & 0 & 0.5\\0 & 1 & 0.5\\0.5 & 0.5 & 1\end{smallmatrix}\right)$ 或 (ii) $\mathcal{MVN}(\left(\begin{smallmatrix}0\\0\\2\end{smallmatrix}\right), \left(\begin{smallmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{smallmatrix}\right)$
$\pi_1(X)$	进入目标区域的抽样概率	$\pi_1(X) = \{1 + \exp(X^\top \eta_1 + 0.6)\}^{-1}$，其中 $\eta_1 = (0.5, 0.3)$;
$\pi_0(X)$	进入辅助区域的抽样概率	$\pi_0(X) = \{1 + \exp(X^\top \eta_0 - 0.4)\}^{-1}$ with $\eta_0 = (-0.5, -0.2)$;
R	抽样指示变量	伯努利分布($\pi(X)$)，其中 $\pi(X) = \frac{\pi_1(X)}{\pi_1(X)+\pi_0(X)}$
A	治疗分配	每个区域内服从伯努利分布(0.5)
β_a	结局模型中的回归系数	$\beta_0 = (0, 2, 2)$; $\beta_1 = (3, 3, 3)$
α_a	未观测 U的系数	$\alpha_0 = \{0.1, 0.5, 1, 1.5\}$; $\alpha_1 = 2\alpha_0$
ε	误差方差	$\varepsilon \in \{0.1, 0.5, 1, 1.5\}$
b_0, b_1	辅助区域中的隐藏偏倚	$b_0 = 6$; $b_1 = 10$
ρ	有偏辅助样本比例	50%

表 S2

Baseline patient characteristics by treatment group across regions. († Median (Q1, Q3); n (%)).

特征	总体 (N = 345†)	辅助区域 GTx-024 3mg (N = 121†)	辅助区域 安慰剂 (N = 155†)	北美 GTx-024 3毫克 (N = 39†)	北美 安慰剂 (N = 30†)
Age	65 (60, 71)	64 (60, 70)	65 (60, 70)	64 (59, 72)	67 (57, 72)
Sex					
F	147 (43%)	50 (41%)	67 (43%)	17（44%）	13（43%）
M	198（57%）	71（59%）	88（57%）	22（56%）	17 (57%)
ECOG基线					
0 – 正常活动；无症状	135 (39%)	43 (36%)	64 (41%)	19 (49%)	9 (30%)
1 – 有症状；完全可行走	210 (61%)	78 (64%)	91 (59%)	20（51%）	21（70%）
化疗					
非紫杉烷类	165（48%）	60（50%）	72（46%）	17（44%）	16（53%）
紫杉烷	180（52%）	61（50%）	83 (54%)	22 (56%)	14 (47%)
癌症分期					
III期	106 (31%)	42 (35%)	44 (28%)	11 (28%)	9 (30%)
IV期	239 (69%)	79 (65%)	111 (72%)	28 (72%)	21 (70%)
组织学					
非鳞状细胞癌	213 (62%)	71 (59%)	96 (62%)	22 (56%)	24 (80%)
鳞状细胞癌	132（38%）	50（41%）	59（38%）	17（44%）	6（20%）
LBM基线	44 (37, 52)	45 (38, 51)	44 (38, 52)	43 (36, 53)	40 (37, 52)
体重减轻标志					
0	32 (46%)	0 (NA%)	0 (NA%)	17 (44%)	15 (50%)
1	37 (54%)	0（不适用%）	0（不适用%）	22（56%）	15（50%）
未知	276	121	155	0	0