arXiv:2601.03480v1 [stat.ME] 7 Jan 2026

通过倾向评分加权的分部分借用幂先验增强对照组，改善临床试验的运行特性

Apu Chandra Das[1*]，Sakib Salam[2], AnindaRoy[3], RakhiChowdhury[4], Antar Chandra Das[5†]，Ashim Chandra Das[6†], 阿尔茨海默病神经影像计划研究人员[7]1*生物统计学系，内布拉斯加大学医学中心，美国内布拉斯加州奥马哈市内布拉斯加大学医学中心984375号，邮编68198。[2, 3]生物统计学部，威斯康星医学院，美国威斯康星州密尔沃基市沃特敦普兰克路8701号，邮编53226。[4]药学系，内布拉斯加大学医学中心，美国内布拉斯加州奥马哈市内布拉斯加大学医学中心984375号，邮编68198。[5,6]软件工程系，达福迪尔国际大学，Daffodil Smart City，孟加拉国达卡萨瓦尔，邮编1216。[7]本文准备过程中使用的数据取自阿尔茨海默病神经影像计划（ADNI）数据库（adni.loni.usc.edu）。因此，ADNI研究人员参与了ADNI的设计和实施和/或提供了数据，但未参与本报告的分析或撰写。ADNI研究人员的完整名单请见：http://adni.loni.usc.edu/wp- content/uploads/how to apply/ADNI Acknowledgement List.pdf。

*通讯作者。电子邮箱：apdas@unmc.edu；共同作者：msalam@mcw.edu；aroy@mcw.edu；racho wdhury@unmc.edu；antar2305341489@diu.edu.bd；ashim2305341488@diu.edu.bd；

†这些作者对本研究的贡献相同。

摘要

借用外部数据可以提高估计效率，但当人群的协变量分布或结局变异性不同时，也可能引入偏倚。为使借用合理，必须在两个数据集之间保持适当的平衡。我们提出倾向评分加权分部借用幂先验（PSW-BPP），将通过倾向评分加权进行的因果协变量调整与灵活的贝叶斯借用方法相结合，在统一框架中应对这些挑战。该方法首先应用倾向评分加权，使外部数据的协变量分布与当前研究的协变量分布相匹配，从而针对共同估计目标，并减少人群异质性造成的混杂。随后，通过分部借用幂先验将加权外部似然纳入贝叶斯模型。该先验允许似然的均值和方差分量分别采用不同的幂参数，从而实现差异化且经过校准的信息借用。此外，我们采用最小合理性指数（mPI）的思路来计算幂参数。与采用单一借用参数的传统幂先验方法相比，这种分别借用的方式能够更稳健地应对先验-数据冲突。我们通过广泛的模拟和一个真实数据示例研究了 PSW-BPP 的运行特性。模拟研究表明，与不借用和固定借用相比，PSW-BPP 能够实现更高效、更稳定的估计，尤其是在协变量中度失衡且存在结局异质性的情况下。该框架为在观察性研究和混合研究设计中使用外部数据进行贝叶斯推断，提供了一项有理有据且可扩展的方法学贡献。

关键词： 倾向评分加权、分部借用幂先验、动态借用、外部对照

1 引言

在临床试验中，随机对照试验（RCTs）被视为药物开发过程中评估治疗效果的金标准 [1]。尽管随机对照试验能够对治疗疗效进行可靠且无偏倚的估计，但在实践中往往面临挑战，包括患者无法及时获得有前景的疗法、试验周期较长、参与者可能接受疗效欠佳的治疗，以及实施上的困难，尤其是在儿科人群中[2]。近几十年来，研究者日益关注将真实世界数据（RWD）纳入药物开发流程，以提高临床试验的效率并确保其符合伦理规范 [3]。一种颇具前景的方法是利用外部临床试验数据来补充新研究的对照组。适当纳入此类外部对照数据，可以减少接受疗效较差治疗的患者人数，缩小对照组规模，并提高试验的整体可行性和伦理平衡 [4]。尽管这项技术可应用于临床试验、遗传学、医疗保健、制造业、环境健康、工程学、经济学和商业等多个领域，但它可能尤其适用于

儿科临床试验，因为儿童及其家长在安排日程方面存在困难，且还需考虑适龄制剂、适口性问题和伦理限制，因此患者招募更具挑战性 [5]。

在利用 RWD 之前，评估当前研究人群与外部研究受试者之间的系统性差异非常重要，以尽量降低潜在偏倚的风险。倾向评分（PS） [6] 是用于减轻此类偏倚的广泛应用的统计工具。除基于 PS 的方法外，研究者还提出了多种贝叶斯方法，将外部信息纳入信息先验。这些方法包括幂先验 [7], 、相容幂先验 [8],、荟萃分析预测（MAP）先验 [9], 、弹性 MAP 先验 [10], 和尺度变换幂先验 [11]。在儿科临床试验中，由于儿科数据有限，借用成人数据十分常见，倾向评分整合贝叶斯方法已得到有效应用。倾向评分分层有助于减少系统性差异，使两个数据集更具可比性 [12, 13]。一项近期研究 [14] 采用倾向评分匹配，在成人受试者与儿科受试者之间进行一对一匹配，随后使用混合先验和幂先验等贝叶斯借用方法估计治疗效应。Wang 等 [15] 和 Lu 等 [16] 将贝叶斯幂先验与倾向评分分层相结合，以推导层特异性后验分布。然而，传统幂先验在整个外部数据集上统一使用单一幂参数，限制了其有选择地借用更相关或质量更高数据部分的灵活性。当外部数据具有异质性时，这可能导致次优借用。为解决这一问题，Baron 等 [17] 近期将 Wang 和 Lu [15, 16] 提出的倾向评分整合幂先验框架扩展为分部借用幂先验，以利用外部数据估计治疗效应。在罕见病或样本量极小的儿科试验中，可能无法形成多个分层。此外，确定每名外部受试者所贡献的信息比例也是合理的。在此类情况下，Li等人[18] 提出将倾向评分加权与贝叶斯幂先验结合使用，无需分层即可改善幂先验的运行特性，从而更有效地扩充对照组。

Wang 等 [15]和 Lu 等 [16] 提出的 PS 整合幂先验方法，需要进行复杂的数值积分，以计算两个数据集之间的重叠区域，随后利用该区域计算幂参数。然而，这两种方法都依赖于预先指定的“要借用的外部受试者名义数量 (A)”，这在实际中未必总是可行。我们从三个方面扩展了以往的方法。首先，我们采用 PS加权 [18] ，而非 PS 分层。与前两种方法相比，我们的加权方法有一个关键优势：实现更简单。其次，我们采用分部分借用幂先验 [19], 而非幂先验，从而能够灵活地通过数据的不同组成部分进行借用。第三，我们引入 Baron 等 [17], 提出的最小合理性指数（mPI）概念来计算幂参数。

我们的加权方案不同于 Lin 等人提出的方法 [14]。Lin 等人 [14] 采用倾向评分匹配来“引入”外部受试者，以扩充当前研究，然后将倾向评分用作受试者特异性幂先验中的幂参数，

而我们则直接通过似然为每个外部对照分配个体特异性权重。与 Li等人 [18], 将权重定义为倾向评分优势比（其值可能超过 1）的方法不同，我们建议使用 1 和倾向评分优势比中的较小值，以确保所有权重都处于 $[0, 1]$ 范围内。

本文其余部分的结构如下。第2 节介绍倾向评分框架、分部借用幂先验以及我们提出的方法。第3 节报告用于评估所提方法性能的模拟设置和结果。真实数据应用的研究结果总结于第4节。最后，第5 节将讨论所提方法及未来研究的潜在方向。

2 方法

2.1 研究设置

我们考虑一项旨在评估研究治疗效应的双臂随机对照试验。除试验数据外，还可从外部对照患者中获得与结局相关的协变量，并可将其纳入分析。主要目标是利用这些外部真实世界数据（RWD），为当前研究中的治疗效应生成无偏且有效的估计。在全文中，我们将外部真实世界对照称为组 e。在试验中，随机分配至对照组的患者构成组 c，而分配至研究治疗组的患者构成组 t。组 c 和组 t 共同构成试验人群。为便于说明，我们聚焦于单个外部数据集，但所提出的策略自然可以扩展到包含多个外部来源的情形。

令 Y_{ti} for $i = 1, \ldots, n_t$, Y_{cj} for $j = 1, \ldots, n_c$ 和 Y_{ek} for $k = 1, \ldots, n_e$ 分别表示研究治疗、同期对照和外部对照样本的连续响应。另令 $\boldsymbol{X}_{ti}$、$\boldsymbol{X}_{cj}$ 和 $\boldsymbol{X}_{ek}$ 表示 p 维协变量向量，其对应的回归系数为 $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$。同时假设 $Y_{ti} \overset{\text{iid}}{\sim} \mathcal{N}(\theta_t, \sigma_t^2)$ $Y_{cj} \overset{\text{iid}}{\sim} \mathcal{N}(\theta_c, \sigma_c^2)$，且 $Y_{ek} \overset{\text{iid}}{\sim} \mathcal{N}(\theta_e, \sigma_e^2)$，其中 θ_t、θ_c 和 θ_e 分别为三组的平均响应，σ_t^2、σ_c^2 和 σ_e^2 分别为三组的方差。治疗效应 $\theta = \theta_t - \theta_c$，即治疗组与对照组平均响应之差，是 $\boldsymbol{\beta}$ 的一部分。我们分别将 $D_t = \{n_t, Y_{ti}, \boldsymbol{X}_{ti}, i = 1, \ldots, n_t\}$、$D_c = \{n_c, Y_{cj}, \boldsymbol{X}_{cj}, j = 1, \ldots, n_c\}$ 和 $D_e = \{n_e, Y_{ek}, \boldsymbol{X}_{ek}, e = 1, \ldots, n_k\}$ 称为研究治疗数据、同期对照数据和外部对照数据。

2.2 倾向评分

倾向评分的概念最初由Rosenbaum和Rubin [6] 提出，作为在观察性研究中进行因果推断的一种方法，此后已广泛应用于各类研究领域 [20–23]。在因伦理、实际操作或经济原因而无法开展随机对照试验的观察性研究中，治疗分配往往受到患者特征的影响，因而可能产生混杂。倾向评分定义为给定观测协变量时接受治疗的概率，是缓解这一问题的一种平衡工具。通过使治疗组和对照组患者

的倾向评分值相近，并采用匹配、分层、加权或协变量调整等方法，可以减少选择偏倚，并使两组达到近似随机对照试验所实现的平衡状态 [24]。这进而提高了基于真实世界数据得出的因果关系估计的有效性和可靠性。基于这一初始概念，我们利用倾向评分平衡当前研究人群与外部研究人群之间的协变量，包括预后因素、基线疾病严重程度、生物标志物和既往治疗。

对于一组基线协变量为 $\boldsymbol{X}$ 的受试者，其倾向评分（PS） e（$\boldsymbol{X}$）可写为

$$e(\boldsymbol{X}) = P(Z = 1 \mid \boldsymbol{X}), \tag{1}$$

其中，若该受试者属于当前研究，则 $Z = 1$；若不属于，则 $Z = 0$。它表示在给定基线协变量 X 的条件下，患者属于当前人群而非外部人群的概率。逻辑回归是最常用的倾向评分估计方法，但也可以采用非参数方法。获得估计评分后，再为每位外部受试者计算一个权重。

2.3 分部借用幂先验（BPP）

Ibrahim 和 Chen [25]，提出的幂先验是一种贝叶斯层次模型，通过明确控制借用程度，将外部对照数据纳入当前研究的分析。它由一个称为幂参数的调节参数决定，该参数作为幂作用于外部数据似然，从而能够整合有信息量的先验知识。这种灵活性使幂先验特别适用于当前样本较小的情形；在这些情形下，利用数据可以增加有效样本量，并提高决策的统计效能。

分部借用幂先验（BPP） [19] 是幂先验的一种扩展，它允许选择性地而非统一地纳入外部数据中的信息。该方法不会对整个数据集应用单一幂参数，而是将参数分解为不同的部分，并为每一部分设定不同的借用程度。这样一来，在外部数据与当前数据相似的部分可以借用更多信息，在二者存在差异的部分则借用较少信息。这种选择性借用提升了外部信息的相关性，从而使统计推断更加准确可靠。

假设两个对照组共享相同的参数，我们通过写出 $\theta_c = \theta_e$ 和 $\sigma_c^2 = \sigma_e^2$来构建幂先验。设 $\bar{Y}_c$ 和 $\bar{Y}_e$ 分别表示当前对照组和外部对照组中响应变量的样本均值，S_c^2 和 S_e^2 分别表示其样本方差。那么，针对 θ_c 和 σ_c^2 的BPP可写为

$$\pi(\theta_c, \sigma_c^2 \mid D_e, \boldsymbol{a}) \propto f(\bar{Y}_e \mid \theta_c, \sigma_c^2)^{a_1} \times f(S_e^2 \mid \sigma_c^2)^{a_2} \times \pi_0(\theta_c, \sigma_c^2), \tag{2}$$

其中，$f(\bar{Y}_e \mid \theta_c, \sigma_c^2)$ 和 $f(S_e^2 \mid \sigma_c^2)$ 是 BPP 中对应于 θ_c和 σ_c^2的两个不同部分，$\pi_0(\theta_c, \sigma_c^2)$ 是针对 θ_c和 σ_c^2的初始先验，而 $\boldsymbol{a} = (a_1, a_2)'(0 \leq a_1, a_2 \leq 1)$ 分别是回归系数 θ_c 和方差 σ^2的幂参数。我们可以将方程（2）补充完整如下：

$$\pi(\theta_c, \sigma_c^2 \mid D_e, \boldsymbol{a}) \propto \left\{ (\sigma_c^2)^{-\frac{1}{2}} \exp\left(-\frac{(\theta_c - \hat{\theta}_e)'(\boldsymbol{X}_e'\boldsymbol{X}_e)(\theta_c - \hat{\theta}_e)}{2\sigma_c^2} \right) \right\}^{a_1} \times \left\{ (S_e^2)^{\frac{n_e-3}{2}} (\sigma^2)^{-\frac{n_e-1}{2}} \exp\left(-\frac{(n_e-1)S_e^2}{2\sigma_c^2} \right) \right\}^{a_2} \left(\frac{1}{\sigma_c^2} \right), \quad (3)$$

其中，$S_e^2 = \frac{1}{n_e-1}(\boldsymbol{Y}_e - \boldsymbol{X}_e\hat{\theta}_e)'(\boldsymbol{Y}_e - \boldsymbol{X}_e\hat{\theta}_e)$、$\pi_0(\theta_c, \sigma_c^2) = 1/\sigma_c^2$，且$\hat{\theta}_e$是仅使用外部对照估计的治疗效应。

Remark 1 方程（3）相较于传统幂先验 a_1，通过 a_2和 [25]，能够更灵活地控制对数据不同分解部分的借用量。若 σ_c^2已知，方程（3）可简化为原始幂先验。

2.4 倾向评分加权的分部分借用幂先验（PSW-BPP）

本节将介绍我们提出的方法。该方法的流程图见图 1。对于未知的治疗效应参数 θ_t 和误差方差 σ_t^2，当前研究治疗组 D_t 的似然可写为

$$\mathcal{L}(\theta_t, \sigma_t^2 \mid D_t) \propto (\sigma_t^2)^{-\frac{n_t}{2}} \exp\left(-\frac{(\theta_t - \hat{\theta}_t)'(\boldsymbol{X}_t'\boldsymbol{X}_t)(\theta_t - \hat{\theta}_t)}{2\sigma_t^2} \right) \exp\left(-\frac{(n_t-1)S_t^2}{2\sigma_t^2} \right), \quad (4)$$

其中 $S_t^2 = \frac{1}{n_t-1}(\boldsymbol{Y}_t - \boldsymbol{X}_t\hat{\theta}_t)'(\boldsymbol{Y}_t - \boldsymbol{X}_t\hat{\theta}_t)$ 且 $\hat{\theta}_t = (\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}\boldsymbol{X}_t'\boldsymbol{Y}_t$。为完善该框架，我们采用标准的非信息先验 $\pi(\theta_t, \sigma_t^2) \propto 1/\sigma_t^2$。将该先验与似然结合后，$(\theta_t, \sigma_t^2)$ 的后验分布为

$$\pi(\theta_t, \sigma_t^2 \mid D_t) \propto (\sigma_t^2)^{-\frac{n_t}{2}-1} \exp\left(-\frac{(\theta_t - \hat{\theta}_t)'(\boldsymbol{X}_t'\boldsymbol{X}_t)(\theta_t - \hat{\theta}_t)}{2\sigma_t^2} \right) \exp\left(-\frac{(n_t-1)S_t^2}{2\sigma_t^2} \right). \quad (5)$$

根据这个联合后验分布，可以直接推导出参数的条件分布。治疗效应的系数，$\theta_t \mid \sigma_t^2, D_t \sim \mathcal{N}\left(\hat{\theta}_t, \sigma_t^2(\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}\right)$ 以及方差，$\sigma_t^2 \mid D_t \sim \text{IG}\left(\frac{n_t-1}{2}, \frac{(n_t-1)S_t^2}{2}\right)$。详细内容见补充材料的 S1 节。

当前研究对照组具有类似的似然，可写为

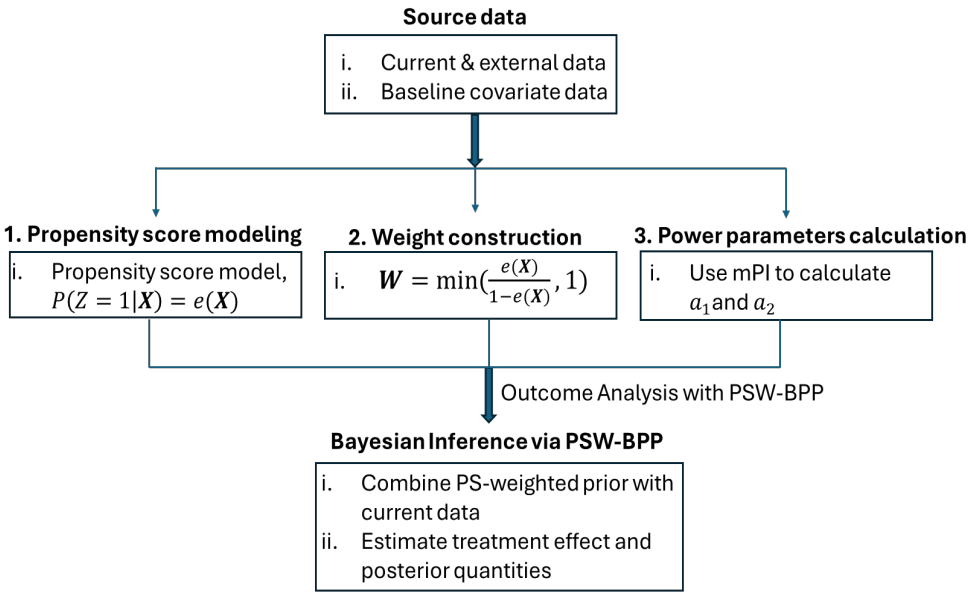

图 1 PSW-BPP 方法示意图

$$\mathcal{L}(\theta_c,\sigma_c^2\mid D_c)\propto(\sigma_c^2)^{-\frac{n_c}{2}}\exp\left(-\frac{(\theta_c-\hat{\theta}_c)'(\boldsymbol{X}_c'\boldsymbol{X}_c)(\theta_c-\hat{\theta}_c)}{2\sigma_c^2}\right)\exp\left(-\frac{(n_c-1)S_c^2}{2\sigma_c^2}\right),\tag{6}$$

其中 $S_c^2=\frac{1}{n_c-1}(\boldsymbol{Y}_c-\boldsymbol{X}_c\hat{\theta}_c)'(\boldsymbol{Y}_c-\boldsymbol{X}_c\hat{\theta}_c)$.

PSW-BPP 可写为

$$\begin{aligned}\pi(\theta_c,\sigma_c^2\mid D_e,\boldsymbol{a})\propto&\left\{(\sigma_c^2)^{-\frac{1}{2}}\exp\left(-\frac{(\theta_c-\hat{\theta}_e^w)'(\boldsymbol{X}_e'\boldsymbol{W}_e\boldsymbol{X}_e)(\theta_c-\hat{\theta}_e^w)}{2\sigma_c^2}\right)\right\}^{a_1}\\&\times\left\{(S_e^{2,w})^{\frac{n_e^*-3}{2}}(\sigma_c^2)^{-\frac{n_e^*-1}{2}}\exp\left(-\frac{(n_e^*-1)S_e^{2,w}}{2\sigma_c^2}\right)\right\}^{a_2}\pi_0(\theta_c,\sigma_c^2),\end{aligned}\tag{7}$$

其中 $\boldsymbol{W}_e=\mathrm{diag}(w_1,\cdots,w_{n_e})$ 是基于倾向评分的权重构成的对角矩阵，$\hat{\theta}_e^w=(\boldsymbol{X}_e'\boldsymbol{W}_e\boldsymbol{X}_e)^{-1}\boldsymbol{X}_e'\boldsymbol{W}_e\boldsymbol{Y}_e$, $S_e^{2,w}=\frac{1}{n_e^*-1}(\boldsymbol{Y}_e-\boldsymbol{X}_e\hat{\theta}_e^w)'\boldsymbol{W}_e(\boldsymbol{Y}_e-\boldsymbol{X}_e\hat{\theta}_e^w)$, $\boldsymbol{a}=(a_1,a_2)'(0\le a_1,a_2\le 1)$ 为幂参数，$n_e^*=\sum_{k=1}^{n_e}w_k$ 是外部对照组中的有效样本量（ESS）。

Li等人 [18] 提出了一种针对外部受试者的加权方案，定义如下

$$w_k=\frac{e(\boldsymbol{X}_{ek})}{1-e(\boldsymbol{X}_{ek})},$$

其中，$e(\boldsymbol{X}_{ek})$ 表示外部受试者的估计倾向评分。然而，直接应用此公式可能会导致权重大于 1，从而

造成后续分析不稳定并降低可解释性。为解决这一问题，我们采用了一种经过修改的加权规则，将其限制在区间 $[0,1]$, 内，该区间

$$w_k = \min\left(1, \frac{e(\boldsymbol{X}_{ek})}{1-e(\boldsymbol{X}_{ek})}\right). \tag{8}$$

此步骤的主要目标是为 [26, 27]外部对照中的每位受试者赋予平衡权重。调整两个对照组之间观测协变量的差异至关重要，从而减轻潜在的混杂偏倚。此重新加权过程有助于构建一个更具可比性的研究人群，进而降低先验-数据冲突的可能性；当外部研究人群与当前研究人群的协变量分布不同时，便可能出现这种冲突。最终，使用平衡权重可使不同人群更为一致，并确保观测到的任何治疗效应较少受到基线特征系统性差异的影响，从而提高分析的有效性。

利用似然和 PSW-BPP，$(\theta_c,\ \sigma_c^2)$ 的联合后验分布可写为

$$\begin{aligned}\pi(\theta_c,\sigma_c^2 \mid D_c, D_e, \boldsymbol{a}) \propto \exp&\left\{-\frac{1}{2\sigma_c^2}\left[(\theta_c-\hat{\theta}_c)'(\boldsymbol{X}_c'\boldsymbol{X}_c)(\theta_c-\hat{\theta}_c)+a_1(\theta_c-\hat{\theta}_e^w)'(\boldsymbol{X}_e'\boldsymbol{W}_e\boldsymbol{X}_e)(\theta_c-\hat{\theta}_e^w)\right]\right\}\\ &\times(\sigma_c^2)^{-\frac{n_c+a_1}{2}}(S_e^{2,w})^{\frac{(n_e^*-3)a_2}{2}}(\sigma_c^2)^{-\frac{(n_e^*-1)a_2}{2}}\\ &\times\exp\left\{-\frac{1}{2\sigma_c^2}\left[(n_c-1)S_c^2+(n_e^*-1)a_2S_e^{2,w}\right]\right\}\left(\frac{1}{\sigma_c^2}\right),\end{aligned} \tag{9}$$

其中，初始先验 $\pi_0(\theta_c,\ \sigma_c^2) \propto 1/\sigma_c^2$为非信息先验。

通过一些代数运算，可以轻松推导出 θ_c 和 σ_c^2 的条件分布。详细内容见补充材料的 S2 节。给定其他变量时，θ_c 的条件后验分布为

$$\theta_c|\sigma_c^2, D_c, D_e, a_1 \sim \mathcal{N}(\tilde{\theta}_c, \Sigma_c), \tag{10}$$

其中

$$\begin{aligned}\tilde{\theta}_c &= (\boldsymbol{X}_c'\boldsymbol{X}_c + a_1\boldsymbol{X}_e'\boldsymbol{W}_e\boldsymbol{X}_e)^{-1}(\boldsymbol{X}_c'\boldsymbol{X}_c\hat{\theta}_c + a_1\boldsymbol{X}_e'\boldsymbol{W}_e\boldsymbol{X}_e\hat{\theta}_e^w),\\ \Sigma_c &= \sigma_c^2(\boldsymbol{X}_c'\boldsymbol{X}_c + a_1\boldsymbol{X}_e'\boldsymbol{W}_e\boldsymbol{X}_e)^{-1}.\end{aligned}$$

类似地，给定其他参数后，σ_c^2 的条件后验分布为

$$\sigma_c^2|D_c, D_e, \boldsymbol{a} \sim \text{IG}\left(\frac{n_c+a_1+(n_e^*-1)a_2}{2}, \frac{(n_c-1)S_c^2+(n_e^*-1)a_2S_e^{2,w}}{2}\right) \tag{11}$$

Remark 2 θ_c 的后验抽样取决于 σ_c^2，而 σ_c^2是根据当前对照组和外部对照所提供的有效样本量估算的。当外部对照组中的受试者少于两名时，σ_c^2 仅根据当前

对照组计算。这项调整会直接影响幂参数的计算。为避免此类情况下不必要的计算复杂度，当外部对照组中的受试者少于两名时，我们将幂参数设为$a_1 = a_2 = 0$。

2.5 幂参数的选择

计算幂参数 $\boldsymbol{a} = (a_1,\ a_2)'$ 是一项颇具挑战的任务。一种方法是将这些参数视为随机变量，并为其指定先验分布，例如 Beta 分布或均匀分布。然而，这种策略往往会带来显著的数学和计算复杂性。也可以将它们固定在不同水平，并进行敏感性分析。作为替代方案，Lu 等 [16]提出利用两组数据之间的重叠系数来计算幂参数，并预先设定一个调节参数，即从外部数据中借用的名义受试者数。为确定重叠系数，Inman 和 Bradley [28] 提出测量两项研究中倾向评分密度曲线的重叠区域。更多详情请参阅 Lu 等 [16]。然而，这种方法需要对两个密度函数之间进行数值积分求解，计算量可能很大。此外，直接指定借用的受试者数可能并不合理。最重要的是，这些方法是为计算层特异性功效参数而开发的，不能直接应用于我们的情况。最近，Baron 等 [17] 提出了一项新准则——最小合理性指数（mPI），该准则具有理论依据，并能为 a_1 和 a_2 提供最优解。

关键思路是衡量外部对照人群与当前对照人群相似的可能性有多大，然后据此设定借用强度。对于治疗效应，令 $\delta = \theta_c - \theta_e$ 表示当前对照组与外部对照组的平均响应之差。零假设$H_0:\ \delta = 0$ 对应于两组在均值方面完全相似的情况，而备择假设 $H_1:\ \delta \neq 0$ 则表示两组存在差异。在标准假设下，δ的后验分布服从一个经过缩放的 t 分布，即 $\frac{\bar{Y}_c - \bar{Y}_e}{S_c/\sqrt{n_c}} \mid (D_c,\ D_e) \sim t_{n_c-1}$，该分布是单峰的，且关于 0 对称。随后，将均值的 mPI 定义为 $\delta = 0$附近的后验概率质量。概率越小，表明相似性越差，因此应降低用于借用均值的幂参数。

对于方差，定义 $\nu = \sigma_c^2/\sigma_e^2$ 为当前对照组与外部对照组的方差比。零假设 $H_0:\ \nu = 1$ 表示两组的变异性相同，而备择假设 $H_1:\ \nu \neq 1$ 则允许方差存在差异。使用参考先验时，样本方差比服从 F分布，$\frac{S_e^2}{S_c^2} \mid (D_c,\ D_e) \sim F_{n_e-1,\ n_c-1}$。$\nu$ 的后验分布为单峰分布，方差的 mPI 对应于 $\nu = 1$的后验可信度。当 $\nu = 1$ 附近的后验概率质量较小时，方差相似度较低，因此应减少对方差分量的借用。建议读者查阅[17]以了解更多详情。

3 模拟研究

3.1 模拟设置

为评估所提出方法的性能，我们开展了大量模拟研究。我们假设当前研究经过了恰当的随机化，从而确保治疗组与当前对照组中的协变量均衡，即它们来自相同的底层分布。我们使用多元正态分布为所有患者模拟相互独立的协变量，共有 p 个协变量，$\boldsymbol{X}_{ti} = (x_{ti1}, \ldots\ldots, x_{tip}) \sim \mathcal{N}_p(\boldsymbol{\mu}_t, \boldsymbol{\Sigma})$，$\boldsymbol{X}_{cj} = (x_{cj1}, \ldots\ldots, x_{cjp}) \sim \mathcal{N}_p(\boldsymbol{\mu}_c, \boldsymbol{\Sigma})$，以及$\boldsymbol{X}_{ek} = (x_{ek1}, \ldots\ldots, x_{ekp}) \sim \mathcal{N}_p(\boldsymbol{\mu}_e, \boldsymbol{\Sigma})$，其中 $i = 1, \ldots\ldots, n_t$、$j = 1, \ldots\ldots, n_c$和 $k = 1, \ldots\ldots, n_e$。结局 Y_{ti}，Y_{cj}、以及 Y_{ek} 通过以下方程生成

$$
\begin{aligned}
\boldsymbol{Y}_{ti} | \boldsymbol{X}_{ti} &= \beta_0 + \boldsymbol{X}'_{ti}\boldsymbol{\beta} + \epsilon_{ti}, \\
\boldsymbol{Y}_{cj} | \boldsymbol{x}_{cj} &= \beta_0 + \boldsymbol{X}'_{cj}\boldsymbol{\beta} + \epsilon_{cj}, \\
\boldsymbol{Y}_{ek} | \boldsymbol{x}_{ek} &= \beta_0 + \boldsymbol{X}'_{ek}\boldsymbol{\beta} + \epsilon_{ek},
\end{aligned}
$$

其中，对于 $i = 1, \cdots, n_t$, $j = 1, \cdots, n_c$和$k = 1, \cdots, n_e$，ϵ_{ti}，$\epsilon_{cj} \sim \mathcal{N}(0, \sigma^2)$和 $\epsilon_{ek} \sim \mathcal{N}(0, \eta^2)$。我们考虑 $n = n_t + n_c = 100$和200, $n_e = 1000$和2000, $p = 4$,$\boldsymbol{\Sigma} = I_4$, $\beta_0 = 0$以及 $\boldsymbol{\beta} = (\theta, 1, 1.5, -1.3)$，其中 $\theta = 2$表示真实治疗效应。

我们在假设 $\boldsymbol{\mu}_t = \boldsymbol{\mu}_c = \boldsymbol{\mu}$的情况下模拟当前研究人群。第一个协变量虽然是连续变量，但在其中位数处二分，以定义治疗指示变量：取值低于中位数的受试者被分配至对照组，而取值大于或等于中位数的受试者被分配至治疗组。相比之下，外部研究中的所有 n_e 受试者均假设属于对照组。需要注意的是，治疗变量通过对第一个协变量进行二分而创建；因此，该协变量原始的连续形式不纳入倾向评分模型和结局模型。不过，其余所有协变量均纳入这两个模型。我们的目标是估计当前研究的治疗效应，同时从外部来源借用最优量的信息。

3.2 评估指标

我们在十二种情景下进行了模拟研究，这些情景涵盖当前数据集与外部对照数据集之间协变量分布和方差结构的差异，以评估所提出的 PSW-BPP方法的性能。所有情景均列于表1。在所有模拟设置中，每种情景均重复 100 次。我们使用预热期 2,000 次迭代后的 5,000 个吉布斯样本，计算每个模拟数据集的所有后验估计。我们计算总体治疗效应的六个后验统计量 θ，以评估所提出模型的性能：(i) 偏倚，表示为 Bias $= \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_b - \theta_b)$，其中 $\hat{\theta}_b$ 表示第 b 个重复数据集的后验均值，θ_b 表示当前组中的真实参数 $b = 1, \cdots, B$；(ii) 绝对偏倚，表示为 ABias $= \frac{1}{B}\sum_{b=1}^{B}|\hat{\theta}_b - \theta_b|$；(iii) 均方根误差，

表示为 RMSE $= \frac{1}{B}\sqrt{\sum_{b=1}^{B}(\hat{\theta}_b - \theta_b)^2}$；(iv) 模拟标准误，表示为 SE $= \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}_b - \frac{1}{B}\sum_{\ell=1}^{B}\hat{\theta}_\ell)^2}$；(v) 95%可信区间的平均宽度（Width）；以及 (vi) 95%可信区间的覆盖概率（CP）。

为了比较我们所提方法的表现，我们还评估了一种替代方法。该方法基于 PSW-BPP 的思路，但不使用 mPI 计算幂参数，而是将其固定为 0、0.5 和 1。设置 $a_1 = a_2 = 0$ 意味着不借用任何信息。通过比较 PSW-BPP 与 FB，我们旨在评估 PS 加权和数据驱动的最优借用如何影响治疗效应估计的准确性和效率。

表1 不同的模拟设置。请注意，每个向量“ –”的第一个元素与治疗变量有关

情景	当前研究		外部研究	
	$\boldsymbol{\mu}$	σ^2	$\boldsymbol{\mu}_e$	η^2
I	(-, 1.2, 1.5, 1.6)	1	(-, 1, 1, 1)	1
II	(-, 1, 1, 1)	1	(-, 1, 1, 1)	1
III	(-, 1.2, 1.5, 1.6)	3	(-, 1, 1, 1)	3
IV	(-, 1, 1, 1)	3	(-, 1, 1, 1)	3
V	(-, 1.2, 1.5, 1.6)	10	(-, 1, 1, 1)	10
VI	(-, 1, 1, 1)	10	(-, 1, 1, 1)	10
VII	(-, 1.2, 1.5, 1.6)	1	(-, 1, 1, 1)	1.5
VIII	(-, 1, 1, 1)	1	(-, 1, 1, 1)	1.5
IX	(-, 1.2, 1.5, 1.6)	3	(-, 1, 1, 1)	4
X	(-, 1, 1, 1)	3	(-, 1, 1, 1)	4
XI	(-, 1.2, 1.5, 1.6)	10	(-, 1, 1, 1)	12
XII	(-, 1, 1, 1)	10	(-, 1, 1, 1)	12

3.3 模拟结果

虽然表2 总结了用于最优借用的 mPI，但表3–6 总结了十二种情景下的模拟结果。我们可以从这些结果中得出若干结论。

表2 显示，在全部十二种情景中，均值分量的 mPI 相对稳定（中位数约为 0.25）；而在存在异质性的情景（情景IX–XII）中，方差分量的 mPI 则大幅下降。这一模式证实，PSW-BPP 能够根据研究之间的异质性有效调整借用强度。

表2 各种场景下100次迭代的 mPI 汇总

情景	$n = 100$				$n = 200$			
	均值的 mPI		方差的 mPI		均值的 mPI		方差的 mPI	
	中位数	IQR	中位数	IQR	中位数	IQR	中位数	IQR
I	0.267	0.235	0.158	0.197	0.215	0.274	0.190	0.233
II	0.275	0.256	0.157	0.197	0.234	0.249	0.190	0.233
III	0.258	0.271	0.160	0.202	0.220	0.279	0.203	0.207
IV	0.280	0.268	0.160	0.202	0.207	0.219	0.203	0.207
V	0.249	0.258	0.151	0.240	0.209	0.234	0.181	0.224
VI	0.254	0.245	0.151	0.240	0.244	0.235	0.181	0.224
VII	0.268	0.233	0.150	0.205	0.216	0.276	0.159	0.214
VIII	0.278	0.256	0.150	0.205	0.233	0.248	0.159	0.214
IX	0.256	0.265	0.091	0.190	0.224	0.272	0.104	0.201
X	0.279	0.266	0.091	0.190	0.211	0.216	0.104	0.201
XI	0.246	0.245	0.105	0.169	0.206	0.244	0.092	0.202
XII	0.251	0.246	0.105	0.169	0.246	0.231	0.092	0.202

Table 3 Simulation results for the overall treatment effect θ across scenarios I–VI ($n = 100$, $n_e = 1000$, and $\sigma^2 = \eta^2$)

Strategy	Scenario	n	n_e	θ	Mean	Bias	ABias	RMSE	SE	Width	CP
PSW-BPP	I	100	1000	2	2.011	−0.011	0.133	0.169	0.169	0.689	0.95
FB ($a_1 = a_2 = 0.0$)		100	1000	2	2.013	−0.013	0.148	0.189	0.189	0.738	0.94
FB ($a_1 = a_2 = 0.5$)		100	1000	2	2.015	−0.015	0.124	0.156	0.156	0.657	0.95
FB ($a_1 = a_2 = 1.0$)		100	1000	2	2.016	−0.016	0.118	0.149	0.149	0.625	0.96
PSW-BPP	II	100	1000	2	2.015	−0.015	0.123	0.157	0.157	0.662	0.96
FB ($a_1 = a_2 = 0.0$)		100	1000	2	2.015	−0.015	0.137	0.177	0.177	0.704	0.93
FB ($a_1 = a_2 = 0.5$)		100	1000	2	2.015	−0.015	0.117	0.148	0.148	0.640	0.96
FB ($a_1 = a_2 = 1.0$)		100	1000	2	2.015	−0.015	0.112	0.140	0.140	0.614	0.97
PSW-BPP	III	100	1000	2	2.014	−0.014	0.233	0.297	0.298	1.193	0.94
FB ($a_1 = a_2 = 0.0$)		100	1000	2	2.023	−0.023	0.256	0.327	0.328	1.278	0.94
FB ($a_1 = a_2 = 0.5$)		100	1000	2	2.026	−0.026	0.214	0.270	0.271	1.138	0.95
FB ($a_1 = a_2 = 1.0$)		100	1000	2	2.027	−0.027	0.204	0.258	0.258	1.083	0.96
PSW-BPP	IV	100	1000	2	2.023	−0.023	0.217	0.277	0.277	1.147	0.95
FB ($a_1 = a_2 = 0.0$)		100	1000	2	2.026	−0.026	0.238	0.306	0.307	1.220	0.93
FB ($a_1 = a_2 = 0.5$)		100	1000	2	2.027	−0.027	0.203	0.256	0.256	1.108	0.96
FB ($a_1 = a_2 = 1.0$)		100	1000	2	2.027	−0.027	0.194	0.243	0.243	1.063	0.97
PSW-BPP	V	100	1000	2	2.023	−0.023	0.434	0.553	0.555	2.179	0.94
FB ($a_1 = a_2 = 0.0$)		100	1000	2	2.041	−0.041	0.467	0.599	0.599	2.334	0.94
FB ($a_1 = a_2 = 0.5$)		100	1000	2	2.047	−0.047	0.391	0.494	0.494	2.078	0.95
FB ($a_1 = a_2 = 1.0$)		100	1000	2	2.050	−0.050	0.373	0.472	0.471	1.977	0.96
PSW-BPP	VI	100	1000	2	2.041	−0.041	0.405	0.517	0.518	2.099	0.95
FB ($a_1 = a_2 = 0.0$)		100	1000	2	2.048	−0.048	0.434	0.560	0.560	2.227	0.93
FB ($a_1 = a_2 = 0.5$)		100	1000	2	2.049	−0.049	0.371	0.467	0.467	2.024	0.96
FB ($a_1 = a_2 = 1.0$)		100	1000	2	2.049	−0.049	0.354	0.044	0.044	1.941	0.97

Table 4 Simulation results for the overall treatment effect θ across scenarios VII–XII ($n = 100$, $n_e = 1000$, and $\sigma^2 \neq \eta^2$)

Strategy	Scenario	n	n_e	θ	Mean	Bias	ABias	RMSE	SE	Width	CP
PSW-BPP	VII	100	1000	2	2.011	−0.011	0.134	0.170	0.170	0.710	0.95
FB ($a_1 = a_2 = 0.0$)		100	1000	2	2.013	−0.013	0.148	0.189	0.189	0.738	0.94
FB ($a_1 = a_2 = 0.5$)		100	1000	2	2.015	−0.015	0.125	0.158	0.158	0.711	0.97
FB ($a_1 = a_2 = 1.0$)		100	1000	2	2.016	−0.016	0.120	0.152	0.152	0.700	0.97
PSW-BPP	VIII	100	1000	2	2.015	−0.015	0.124	0.157	0.157	0.682	0.96
FB ($a_1 = a_2 = 0.0$)		100	1000	2	2.015	−0.015	0.137	0.177	0.177	0.704	0.93
FB ($a_1 = a_2 = 0.5$)		100	1000	2	2.015	−0.015	0.118	0.148	0.148	0.692	0.98
FB ($a_1 = a_2 = 1.0$)		100	1000	2	2.015	−0.015	0.113	0.141	0.141	0.687	0.98
PSW-BPP	IX	100	1000	2	2.014	−0.014	0.233	0.298	0.300	1.213	0.93
FB ($a_1 = a_2 = 0.0$)		100	1000	2	2.023	−0.023	0.256	0.327	0.328	1.278	0.94
FB ($a_1 = a_2 = 0.5$)		100	1000	2	2.026	−0.026	0.216	0.273	0.273	1.200	0.96
FB ($a_1 = a_2 = 1.0$)		100	1000	2	2.028	−0.028	0.206	0.262	0.262	1.171	0.97
PSW-BPP	X	100	1000	2	2.023	−0.023	0.217	0.277	0.278	1.166	0.96
FB ($a_1 = a_2 = 0.0$)		100	1000	2	2.026	−0.026	0.238	0.306	0.307	1.220	0.93
FB ($a_1 = a_2 = 0.5$)		100	1000	2	2.026	−0.027	0.204	0.256	0.256	1.170	0.98
FB ($a_1 = a_2 = 1.0$)		100	1000	2	2.027	−0.027	0.195	0.244	0.244	1.149	0.98
PSW-BPP	XI	100	1000	2	2.023	−0.023	0.435	0.554	0.556	2.196	0.94
FB ($a_1 = a_2 = 0.0$)		100	1000	2	2.041	−0.041	0.467	0.597	0.599	2.334	0.94
FB ($a_1 = a_2 = 0.5$)		100	1000	2	2.048	−0.048	0.393	0.496	0.496	2.147	0.96
FB ($a_1 = a_2 = 1.0$)		100	1000	2	2.050	−0.050	0.375	0.475	0.475	2.075	0.97
PSW-BPP	XII	100	1000	2	2.041	−0.041	0.405	0.518	0.519	2.116	0.95
FB ($a_1 = a_2 = 0.0$)		100	1000	2	2.048	−0.048	0.434	0.560	0.560	2.227	0.93
FB ($a_1 = a_2 = 0.5$)		100	1000	2	2.048	−0.048	0.372	0.468	0.468	2.091	0.97
FB ($a_1 = a_2 = 1.0$)		100	1000	2	2.049	−0.049	0.355	0.445	0.445	2.037	0.98

Table 5 Simulation results for the overall treatment effect θ across scenarios I–VI ($n = 200$, $n_e = 2000$, and $\sigma^2 = \eta^2$)

Strategy	Scenario	n	n_e	θ	Mean	Bias	ABias	RMSE	SE	Width	CP
PSW-BPP	I	200	2000	2	1.999	0.001	0.090	0.113	0.114	0.492	0.95
FB ($a_1 = a_2 = 0.0$)		200	2000	2	2.006	−0.006	0.099	0.124	0.125	0.524	0.94
FB ($a_1 = a_2 = 0.5$)		200	2000	2	2.002	−0.002	0.085	0.105	0.105	0.465	0.97
FB ($a_1 = a_2 = 1.0$)		200	2000	2	2.001	−0.001	0.084	0.102	0.103	0.442	0.98
PSW-BPP	II	200	2000	2	2.007	−0.007	0.090	0.111	0.112	0.474	0.96
FB ($a_1 = a_2 = 0.0$)		200	2000	2	2.007	−0.007	0.097	0.120	0.121	0.500	0.97
FB ($a_1 = a_2 = 0.5$)		200	2000	2	2.004	−0.004	0.083	0.104	0.104	0.454	0.97
FB ($a_1 = a_2 = 1.0$)		200	2000	2	2.003	−0.003	0.080	0.101	0.101	0.435	0.97
PSW-BPP	III	200	2000	2	1.998	0.002	0.158	0.198	0.199	0.851	0.95
FB ($a_1 = a_2 = 0.0$)		200	2000	2	2.011	−0.011	0.173	0.215	0.216	0.907	0.94
FB ($a_1 = a_2 = 0.5$)		200	2000	2	2.003	−0.003	0.147	0.181	0.182	0.805	0.97
FB ($a_1 = a_2 = 1.0$)		200	2000	2	2.001	−0.001	0.145	0.177	0.178	0.765	0.98
PSW-BPP	IV	200	2000	2	2.011	−0.011	0.156	0.194	0.195	0.822	0.96
FB ($a_1 = a_2 = 0.0$)		200	2000	2	2.012	−0.012	0.168	0.208	0.209	0.866	0.97
FB ($a_1 = a_2 = 0.5$)		200	2000	2	2.007	−0.007	0.144	0.179	0.180	0.789	0.97
FB ($a_1 = a_2 = 1.0$)		200	2000	2	2.005	−0.005	0.139	0.174	0.175	0.754	0.97
PSW-BPP	V	200	2000	2	1.998	0.002	0.291	0.365	0.367	1.553	0.95
FB ($a_1 = a_2 = 0.0$)		200	2000	2	2.019	−0.019	0.316	0.392	0.394	1.655	0.94
FB ($a_1 = a_2 = 0.5$)		200	2000	2	2.006	−0.006	0.268	0.331	0.332	1.470	0.97
FB ($a_1 = a_2 = 1.0$)		200	2000	2	2.001	−0.001	0.264	0.323	0.325	1.397	0.98
PSW-BPP	VI	200	2000	2	2.017	−0.017	0.289	0.355	0.357	1.500	0.96
FB ($a_1 = a_2 = 0.0$)		200	2000	2	2.021	−0.021	0.307	0.380	0.381	1.581	0.97
FB ($a_1 = a_2 = 0.5$)		200	2000	2	2.012	−0.012	0.263	0.327	0.329	1.435	0.97
FB ($a_1 = a_2 = 1.0$)		200	2000	2	2.001	−0.001	0.254	0.318	0.319	1.377	0.97

Table 6 Simulation results for the overall treatment effect θ across scenarios VII–XII ($n = 200$, $n_e = 2000$, and $\sigma^2 \neq \eta^2$)

Strategy	Scenario	n	n_e	θ	Mean	Bias	ABias	RMSE	SE	Width	CP
PSW-BPP	VII	200	2000	2	1.999	0.001	0.091	0.114	0.114	0.508	0.97
FB ($a_1 = a_2 = 0.0$)		200	2000	2	2.006	−0.006	0.100	0.124	0.125	0.523	0.94
FB ($a_1 = a_2 = 0.5$)		200	2000	2	2.001	−0.001	0.085	0.105	0.106	0.502	0.98
FB ($a_1 = a_2 = 1.0$)		200	2000	2	1.999	0.001	0.085	0.104	0.104	0.494	0.98
PSW-BPP	VIII	200	2000	2	2.007	−0.007	0.090	0.111	0.112	0.490	0.96
FB ($a_1 = a_2 = 0.0$)		200	2000	2	2.007	−0.007	0.097	0.121	0.121	0.500	0.97
FB ($a_1 = a_2 = 0.5$)		200	2000	2	2.004	−0.004	0.083	0.104	0.104	0.490	0.98
FB ($a_1 = a_2 = 1.0$)		200	2000	2	2.002	−0.002	0.081	0.101	101	0.487	0.98
PSW-BPP	IX	200	2000	2	1.998	0.002	0.159	0.199	0.200	00.867	0.97
FB ($a_1 = a_2 = 0.0$)		200	2000	2	2.011	−0.011	0.173	0.215	0.216	0.907	0.94
FB ($a_1 = a_2 = 0.5$)		200	2000	2	2.003	−0.003	0.148	0.182	0.183	0.848	0.98
FB ($a_1 = a_2 = 1.0$)		200	2000	2	1.999	0.001	0.147	0.179	0.180	0.827	0.98
PSW-BPP	X	200	2000	2	2.011	−0.011	0.158	0.194	0.195	0.837	0.96
FB ($a_1 = a_2 = 0.0$)		200	2000	2	2.012	−0.012	0.168	0.208	0.209	0.866	0.97
FB ($a_1 = a_2 = 0.5$)		200	2000	2	2.006	−0.006	0.144	0.180	0.180	0.829	0.98
FB ($a_1 = a_2 = 1.0$)		200	2000	2	2.004	−0.004	0.140	0.175	0.175	0.814	0.98
PSW-BPP	XI	200	2000	2	1.997	0.003	0.292	0.365	0.367	1.566	0.96
FB ($a_1 = a_2 = 0.0$)		200	2000	2	2.019	−0.019	0.316	0.392	0.394	1.655	0.94
FB ($a_1 = a_2 = 0.5$)		200	2000	2	2.005	−0.005	0.269	0.332	0.333	1.518	0.98
FB ($a_1 = a_2 = 1.0$)		200	2000	2	2.001	−0.001	0.266	0.325	0.327	1.465	0.98
PSW-BPP	XII	200	2000	2	2.017	−0.017	0.289	0.355	0.357	1.513	0.96
FB ($a_1 = a_2 = 0.0$)		200	2000	2	2.021	−0.021	0.307	0.380	0.381	1.581	0.97
FB ($a_1 = a_2 = 0.5$)		200	2000	2	2.012	−0.012	0.264	0.328	0.329	1.482	0.97
FB ($a_1 = a_2 = 1.0$)		200	2000	2	2.008	−0.008	0.255	0.318	0.320	1.444	0.98

表3 和5 展示了方差齐性（$\sigma^2 = \eta^2$时，针对不同样本量的模拟情景I–VI。在这些设定下，PSW-BPP在所有情景中对总体治疗效应 θ 的估计几乎无偏。偏倚和均方根误差均较小，95%可信区间达到了名义覆盖水平。与未利用信息（$a_1 = a_2 = 0$）时的FB方法相比，PSW-BPP的均方根误差更小、区间宽度更窄，同时覆盖概率相当或略高。然而，在利用更多信息（$a_1 = a_2 \geq 0.5$）时，FB在所有情景中的后验分位数均较低，这可能是由于过度借用所致。这表明，当当前研究与外部研究存在差异时，固定借用可能导致外部数据产生过大的影响；而PSW-BPP的自适应特性能够有效控制借用强度，从而获得更可靠的推断。其次，结果表明，当外部研究与当前研究之间存在异质性（如情景I、III和V）时，与相应的齐性情景（II、IV和VI）相比，绝对偏倚、均方根误差、标准误和区间宽度均有所增加。这一发现说明，两项研究越可比，借用信息越有效。第三，随着当前研究样本量 n从100增加到200，且外部研究样本量 n_e 从1000增加到2000，所有方法的绝对偏倚、均方根误差和标准误均有所降低，区间宽度也变窄。无论样本量如何变化，PSW-BPP均持续展现出稳定的估计准确性和可靠的不确定性量化，证实了在有更大的外部数据集可供借用时，该方法具有良好的可扩展性。这些发现表明，通过BPP纳入外部信息，能够获得比不借用时更准确、更一致的估计。尤其是，覆盖稳定性的提升表明，与不借用方法相比，受控借用可以通过降低变异性和减轻偏倚来提高推断的可靠性。最后，数据驱动的mPI机制和倾向评分加权使PSW-BPP能够动态确定借用强度，而非依赖固定的名义借用样本量，从而实现更稳健、更具自适应性的推断。

表 4 和 6展示了两个数据集之间的方差存在异质性时的结果（$\sigma^2 \neq \eta^2$；情景VII–XII）。如表 2所示，PSW-BPP 自动减少了对外部方差分量的借用，其方差–mPI值较低。这种自适应降权避免了可信区间过窄，并使覆盖率维持在接近名义95%的水平。与FB相比，PSW-BPP提供了稳定的推断、适中的区间宽度和最小的偏倚，凸显了其对异质性外部信息的稳健性。

总体而言，这些结果表明，当当前对照组与外部对照具有同质性时，PS
W-BPP 能够高效地纳入外部信息，并在出现异质性时自动降低借用强度。所提出的方法在广泛的数据生成情景和样本量下，均能提供准确、高效且稳健的治疗效应估计 θ 。

4 案例研究

尽管所提出的方法旨在通过减少所需患者数量，使临床试验更加符合伦理且更高效，但我们未能在公共领域找到合适的个体层面数据。因此，我们通过重新整理一项回顾性、非随机研究的数据，使其模拟随机对照试验（RCT）情景，以展示所提出方法的应用。我们使用阿尔茨海默病研究领域广泛认可的数据库阿尔茨海默病神经影像学计划（ADNI），并据此构建准随机暴露–对照试验和外部对照组。尽管这并不是所提出方法的完美案例研究，但足以达到方法学演示的目的。

被视为评估抗痴呆治疗效果的金标准 [29]，ADAS13由一系列13项认知任务组成，总分定义为各项任务错误数之和；分数越高，表示认知表现越差。我们将暴露效应大小定义为基线与第52周ADAS13评分之差，正差值表示受试者的认知功能有所改善。分析中纳入的基线协变量包括年龄、性别、RAVLT（雷伊听觉词语学习测验）、APOE4状态和MMSE（简易精神状态检查）评分。虽然有五种基线诊断类别可供选择，但我们仅纳入最相关的三类，即认知正常（CN）、晚期轻度认知障碍（LMCI）和阿尔茨海默病（AD），以定义暴露组与对照组变量。具体而言，CN为对照组，LMCI和AD合并构成暴露组。APOE4状态按0、1或2个等位基因分类，MMSE评分范围为18至30。RAVLT指标包括三个子项——RAVLT 即时回忆、学习和遗忘，我们对这三项取平均值，构建单一的RAVLT综合变量。主要数据集包括按入组时期划分的三个队列。我们将ADNI1视为外部数据，将ADNI2和ADNI3视为当前数据，因此当前研究包含531名受试者，外部研究队列包含710名受试者。本分析的主要目标是在借用外部数据中的信息的同时，估计暴露效应大小。基线协变量汇总见表7。

表 7 当前研究与外部研究人群的基线特征

特征	**当前研究** $n = 531$	**外部研究** $n = 710$	p**-值**
暴露（%）	63.65	70.99	0.007
男性（%）	55.18	57.75	0.398
年龄（均值/标准差）	73.19（7.36）	75.27（6.73）	< 0.001
基线听觉词语学习测验（均值/标准差）	14.42 (5.03)	13.73 (4.44)	0.012
APOE4 0 (/%)	50.65	51.41	0.500
1 (/%)	36.34	37.75	
基线简易精神状态检查（均值/标准差）	26.95 (2.93)	26.88 (2.58)	0.629

请注意，分类变量的 p 值采用卡方检验计算，连续变量则采用双样本 t 检验计算。

我们采用所提出的PSW-BPP方法估计暴露效应大小。假设设定如下： H_0 ： $\theta \leq 0$ 对 H_a ： $\theta > 0$，其中 θ 表示暴露效应，定义为暴露组与对照组ADAS13评分变化均值之差。 θ 越大，表示认知改善程度越高。假设当 $\theta > 0$ 的后验概率超过预先设定的0.975阈值时，拒绝零假设。

倾向评分通过逻辑回归模型估计，并为外部队列中的个体赋予个体特异性权重。这些权重被纳入 BPP框架，使外部数据能够按照其与当前研究的相似程度成比例地贡献信息。采用 PSW-BPP方法，估计的暴露效应为 $\theta = 0.166$，95%可信区间为（0.039,0.309），表明该效应在统计学上显著为正。这一结果提示，暴露与 ADAS13评分降低相关，反映出认知表现有所改善。

$\theta > 0$ 的后验概率为99.3%，达到预先设定的研究成功标准。在此框架下，外部数据贡献的有效样本量为153.66，体现了 PSW-BPP方法所实现的部分借用。仅关注当前研究而不借用外部数据时，$\theta = 0.175$ 的95%可信区间为（0.039,0.326），表明纳入外部数据并未实质性改变点估计。然而，借用外部数据缩窄了可信区间，从而使暴露效应的估计更加精确、稳健。两个幂参数估计值分别为$a_1 = 0.482$ 和 $a_2 = 0.215$。因此，从均值部分借用的信息多于从方差部分借用的信息。

总而言之，PSW-BPP 方法能够以合乎伦理且高效的方式纳入外部信息，同时维护当前研究推断的完整性。

5 讨论

本文提出了一个统一框架 PSW-BPP，通过借用外部对照数据中的信息来扩充随机对照试验的对照组。该方法将倾向评分加权与 BPP 相结合，并采用数据自适应幂参数，分别调节结局模型中均值和方差分量所借用的信息量。这种结合实现了灵活的分量级借用，可反映当前对照数据与外部对照数据之间的兼容程度，同时对关注的治疗效应保持有效且高效的推断。

所提出的方法相较于 Li等人的方法具有多项方法学优势。 [18]首先，我们对倾向评分加权方案进行了修改，以确保所得权重被限制在区间 $[0, 1]$，内，从而提高数值稳定性，并防止个别外部观测产生过度影响。其次，我们没有采用仅含一个全局幂参数的传统幂先验，而是使用包含两个不同幂参数的 BPP，以实现对信息的选择性和分量级借用。当结局模型的均值和方差分量在当前数据源与外部数据源之间呈现不同程度的异质性时，这种设定尤其有利

。第三，我们利用 mPI 的概念指导幂参数的校准，提供一种数据驱动机制，以降低过度借用或借用不足的风险，并增强所得推断的稳健性。

模拟研究展示了PSW-BPP相较于其他借用方法的几项关键优势。在广泛的数据生成情景中，包括协变量分布、残差方差和样本量各不相同的情景，PSW-BPP始终能够对总体治疗效应给出近乎无偏的估计，并进行恰当的不确定性量化。与固定借用方法相比，PSW-BPP的均方根误差更低，且生成的可信区间更窄。PSW-BPP的另一项关键优势在于，当研究之间存在差异时，它能够自适应地降低外部信息的权重。当当前对照组与外部对照组之间存在方差或协变量分布异质性时，由mPI驱动的借用机制会自动降低借用程度，从而避免对不确定性作出过于乐观的估计。这种自适应行为有别于传统的固定幂先验及相关方法；即使数据源之间存在部分不相容，这些方法仍可能继续大量借用，从而导致覆盖不足 [8, 25]。

与其他借用策略的比较进一步凸显了PSW-BPP的优势。与预先指定幂参数的固定借用先验相比，PSW-BPP无需主观调整，并降低了对先验设定错误的敏感性。不同于基于分层的方法 [15–17] ，后者依赖于固定的名义借用样本量，PSW-BPP允许在各模型组成部分之间以完全由数据驱动的方式灵活调整借用程度。在几乎所有模拟情景中，所提出的方法在保持覆盖率更加稳定的同时，取得了相当或更高的效率。此外，随着样本量增加，PSW-BPP仍能有效扩展；当前数据集和外部数据集规模增大时，它继续提供准确的点估计和校准良好的区间。这种可扩展性对于当代临床研究尤为重要，因为此类研究越来越多地利用包括真实世界证据在内的大型外部数据源 [30]。

尽管 PSW-BPP 具有若干方法学优势，但也存在一些局限。首先，所提出的框架假设所有相关混杂因素均已观测到，并且在倾向评分模型中得到充分捕捉。与所有基于倾向评分的方法一样，未测量混杂可能会损害加权似然的有效性，进而影响借用机制 [6, 31]。其次，当前实现主要针对具有高斯误差的连续结局；若扩展至广义线性模型和生存结局，将进一步拓宽该方法的适用范围。第三，虽然 mPI 提供了一个可解释且有效的相容性度量，但也可以探索其他差异度量或层级扩展，以进一步增强稳健性。

未来的研究可以将 PSW-BPP 扩展到更复杂的数据设置，包括二分类结局和生存结局、纵向响应变量，以及多层或聚类数据结构。尽管所提出的方法针对单个外部数据源开发，但通过纳入多个外部数据集，可以自然地对其进行扩展

允许使用数据集特异性加权和借用参数。此外，采用基于灵活机器学习的倾向评分模型，可能会进一步改善高维设置下的协变量平衡。另外，还可以在倾向评分模型中针对高维数据进行变量选择。最后，进一步深入研究 mPI 的理论性质，尤其是在模型错设条件下的理论性质，将有助于深入了解所提出借用机制的稳健性和运行特性。

补充信息

在线补充材料包括两个部分：(i) S1（治疗组推断）和 (ii) S2（对照组推断）。

致谢

本项目的数据收集和共享由阿尔茨海默病神经影像计划（ADNI）（美国国立卫生研究院拨款编号 U01 AG024904）和国防部 ADNI（美国国防部资助编号 W81XWH-12-2-0012）提供资金支持。ADNI 由美国国家老龄研究所、美国国家生物医学成像与生物工程研究所提供资助，并得到以下机构的慷慨捐助：艾伯维、阿尔茨海默病协会；阿尔茨海默病药物发现基金会；Araclon Biotech；BioClinica, Inc.；渤健；百时美施贵宝公司；CereSpir, Inc.；Cogstate；卫材公司；Elan Pharmaceuticals, Inc.；礼来公司；EuroImmun；F. Hoffmann-La Roche Ltd 及其关联公司基因泰克公司；富士瑞必欧；通用电气医疗；IXICO Ltd.；Janssen AlzheimerImmunotherapyResearch& Development, LLC.；强生制药研发有限责任公司；Lumos- ity；灵北；默克公司；Meso Scale Diagnostics, LLC.；NeuroRx Research；Neurotrack Technologies；诺华制药公司；辉瑞公司；Piramal Imaging；施维雅；武田药品工业株式会社；以及 Transition Therapeutics。加拿大卫生研究院为支持加拿大 ADNI 临床研究中心提供资金。私人部门的捐助由美国国立卫生研究院基金会促成（www.fnih.org）。项目受资助机构为北加州研究与教育研究所，研究由南加州大学阿尔茨海默病治疗研究所协调。ADNI 数据由南加州大学神经影像实验室发布。

声明

- 经费支持：不适用。
- 利益冲突/竞争性利益：作者声明不存在潜在利益冲突。

- 伦理审批与参与同意：不适用。
- 数据可用性：可向ADNI 申请获取数据。
- 代码可用性：所有代码均使用统计软件 R 4.2.2 编写，可从GitHub下载。

参考文献

[1] Bakker, E., Plueschke, K., Jonker, C.J., Kurz, X., Starokozhko, V., Mol, P.G.：真实世界证据对欧洲药品管理局监管决策的贡献。《临床药理学与治疗学》113(1), 135–151 (2023) https://doi.org/10.1002/cpt.2766[2] Purpura, C.A., Garry, E.M., Honig, N., Case, A., Rassen, J.A.：真实世界证据在美国食品药品监督管理局批准的新药和生物制品许可申请中的作用。《临床药理学与治疗学》111(1), 135–144 (2022) https://doi.org/ 10.1002/cpt.2474[3] Kidwell, K.M., Roychoudhury, S., Wendelberger, B., Scott, J., Moroz, T., Yin, S., Majumder, M., Zhong, J., Huml, R.A., Miller, V.：贝叶斯方法在加速罕见病药物开发中的应用：范围与障碍。《孤儿药杂志》17(1), 186 (2022) https://doi.org/10.1186/s13023-022-02342-5[4] Ghadessi, M., Tang, R., Zhou, J., Liu, R., Wang, C., Toyoizumi, K., Mei, C., Zhang, L., Deng, C., Beckman, R.A.：临床试验中使用历史对照的路线图——由药品信息协会自适应设计科学工作组（dia-adswg）编制。《孤儿药杂志》15(1), 69 (2020) https: //doi.org/10.1186/s13023-020-1332-x[5] Boat, T.F., Field, M.J.：《儿童安全有效的药物：根据《儿童最佳药品法》和《儿科研究平等法》开展的儿科研究》(2012) https://doi.org/10.17226/13311[6] Rosenbaum, P.R., Rubin, D.B.：倾向评分在观察性研究因果效应分析中的核心作用。《生物计量学》70(1), 41–55 (1983) https: //doi.org/10.1093/biomet/70.1.41[7] Ibrahim, J.G., Chen, M.-H., Gwon, Y., Chen, F.：幂先验：理论与应用。《医学统计学》34(28), 3724–3749 (2015) https://doi.org/10. 1002/sim.6728[8] Hobbs, B.P., Carlin, B.P., Mandrekar, S.J., Sargent, D.J.：用于临床试验中自适应纳入历史信息的层级相容先验与幂先验模型。《生物统计学》67(3), 1047–1056 (2011) https://doi.org/10.1111/j.1541-0420.2011.01564.x[9]Schmidli, H., Gsteiger, S., Roychoudhury, S., O’Hagan, A., Spiegelhalter, D., Neuenschwander, B.：包含历史对照信息的临床试验中的稳健的荟萃分析预测先验。《生物统计学》70(4), 1023–1032 (2014) https: //doi.org/10.1111/biom.12242

[10] Zhang, W., Pan, Z., Yuan, Y.：用于动态借用历史数据中的信息的弹性荟萃分析预测先验，并应用于生物类似药临床试验。Contemporary Clinical Trials 110, 106559 (2021) https://doi.org/ 10.1016/j.cct.2021.106559[11] Alt, E.M., Nifong, B., Chen, X., Psioda, M.A., Ibrahim, J.G.：用于采用不同结局模型的历史数据的尺度变换幂先验。《医学统计学》42(1), 1–14 (2023) https://doi.org/10.1002/sim.9598[12] Lunceford, J.K., Davidian, M.：在因果治疗效应估计中通过倾向评分进行分层和加权：一项比较研究。《医学统计学》23(19), 2937–2960 (2004) https://doi.org/10.1002/sim.1903[13]Rosenbaum, P.R., Rubin, D.B.：利用倾向评分亚分类减少观察性研究中的偏倚。《美国统计学会杂志》79(387), 516–524 (1984) https://doi.org/10.1080/01621459.1984. 10478078[14] Lin, J., Gamalo-Siebers, M., Tiwari, R.：用于贝叶斯增强对照设计的基于倾向评分的先验。《药物统计学》18(2), 223–238 (2019) https://doi.org/10.1002/pst.1918 [15] Wang, C., Li, H., Chen, W.-C., Lu, N., Tiwari, R., Xu, Y., Yue, L.Q.：用于在单臂临床研究中纳入真实世界证据的倾向评分整合幂先验方法。生物制药统计学杂志 29(5), 731–748 (2019) https://doi.org/10.1080/10543406.2019.1657133[16] Lu, N., Wang, C., Chen, W.-C., Li, H., Song, C., Tiwari, R., Xu, Y., Yue, L.Q.：通过纳入多个外部数据源，使用倾向评分整合幂先验方法增强随机对照试验的对照组。生物制药统计学杂志 32(1), 158–169 (2022) https://doi.org/ 10.1080/10543406.2021.1998098[17] Baron, E., Lin, M., Zhu, J., Tang, R., Chen, M.-H., Initiative*, A.D.N.：增强随机对照试验：一种用于借用外部对照数据的贝叶斯分而治之方法。生物科学统计学，1–26 (2024) https: // doi.org/10.1007/s12561-024-09465-2[18] Li, H., Chen, W.-C., Wang, C., Lu, N., Song, C., Tiwari, R., Alexander, G., Xu, Y., Yue, L.Q.：使用倾向评分加权，在利用外部数据增强传统临床研究时改善幂先验的运行特性。《药物统计学》24(2), 2471 (2025) https://doi.org/10. 1002/pst.2471[19] Yuan, W., Chen, M.-H., Zhong, J.：优效性试验的贝叶斯设计：方法与应用。生物制药研究统计学 14(4), 433–443 (2022) https://doi.org/10.1080/19466315.2022.2090429

[20] McCaffrey, D.F., Griffin, B.A., Almirall, D., Slaughter, M.E., Ramchand, R., Burgette, L.F.：使用广义提升模型估计多种治疗的倾向评分教程。《医学统计学》32(19), 3388–3414 (2013) https://doi.org/10.1002/sim.5753[21]Pan, W., Bai, H.：因果推断的倾向评分方法：概述。Behaviormetrika 45(2), 317–334 (2018) https://doi.org/10.1007/ s41237-018-0058-8[22] Austin, P.C., Yu, A.Y.X., Vyas, M.V., Kapral, M.K.：在神经病学临床研究中应用倾向评分方法。《神经病学》97(18), 856–863(2021) https: //doi.org/10.1212/WNL.000000000001277[23] D'Agostino Jr, R.B.：倾向评分在心血管研究中的应用。《循环》115(17), 2340–2343 (2007) https://doi.org/10.1161/CIRCULATIONAHA.105. 59495[24] Akmal, S.R., Beier, M.A., August, D.A.：倾向评分方法在外科肿瘤学文献中的应用。《外科肿瘤学》42, 101776 (2022) https://doi.org/10. 1016/j.suronc.2022.101776[25] Ibrahim, J.G., Chen, M.-H.：回归模型的幂先验分布。Statistical Science 15(1), 46–60 (2000) https://doi.org/10.1214/ss/1009212673[26]Li, H., Wang, C., Chen, W.-C., Lu, N., Song, C., Tiwari, R., Xu, Y., Yue, L.Q.：观察性研究中的估计目标：ICH E9 (R1)之外的一些考量。《药物统计学》21(5), 835–844 (2022) https://doi.org/10.1002/ pst.2196[27] Li, F., Morgan, K.L., Zaslavsky, A.M.：通过倾向评分加权平衡协变量。《美国统计学会杂志》113(521), 390–400 (2018) https://doi.org/10.1080/01621459.2016.1260466[28] Inman, H.F., Bradley Jr, E.L.：重叠系数作为概率分布之间一致性的度量，以及两个正态密度重叠部分的点估计。《统计通讯：理论与方法》18(10), 3851–3874 (1989) https://doi.org/03610928908830127[29] Kueper, J.K., Speechley, M., Montero-Odasso, M.：阿尔茨海默病评估量表认知分量表（ADAS-Cog）：痴呆前人群中的修订与响应性。叙述性综述。《阿尔茨海默病杂志》63(2), 423–444 (2018) https://doi.org/10.3233/JAD-170991[30] Hern´an, M.A., Robins, J.M.：随机试验不可用时使用大数据模拟目标试验。《美国流行病学杂志》183(8), 758– 764 (2016) https://doi.org/10.1093/aje/kwv254

[31] Stuart, E.A.：因果推断的匹配方法：回顾与展望。Statistical science：数学统计学会的评论期刊**25**(1), 1（2010）https://doi.org/10.1214/09-STS313