针对异质性总体的倾向性评分加权方法比较研究*

王旭晴[1#]　胡　禹[1#]　秦国友[2△]

【提　要】 **目的**　探讨在观察性研究中，五种倾向性评分加权方法估计异质性总体平均处理效应的性能。**方法**　采用Monte Carlo模拟方法，设置了六种异质性场景以及三种倾向性评分分布重叠程度，生成相应数据集，再将稳定的逆概率加权、修剪的逆概率加权、重叠加权、经验似然加权以及协变量平衡倾向性评分加权估计器分别应用于估计平均处理效应，并根据相对偏差和相对均方根误差两个指标来衡量各估计器的性能。**结果**　在异质性总体中，模型正确指定的情况下，推断目标为整个人群的估计器相对偏差明显小于其他估计器，其中协变量平衡倾向性评分加权估计器表现最佳，其相对均方根误差始终最大。**结论**　在估计异质性总体的平均处理效应时，重叠加权方法可能存在较大偏倚，协变量平衡倾向性评分加权方法在准确性和效率等方面均相对最优。

【关键词】 倾向性评分　平均处理效应　异质性总体　加权
【中图分类号】 R311　　【文献标志码】 A　　**DOI**　10.11783/j.issn.1002-3674.2026.05.013

近年来，估计平均处理效应（average treatment effect，ATE）以揭示处理变量与结果变量间的因果关系受到广泛关注。在观察性研究中，由于处理分配缺乏随机性，ATE的估计往往存在混杂偏倚[1-2]。倾向性评分（propensity score，PS）加权法是调整混杂因素的常用手段，目前已提出了多种倾向性评分加权方法，如逆概率加权、重叠加权、经验似然加权等[3-6]。以往研究多在同质总体中评估比较这些方法[7-10]，然而在实际中，个体对处理的反应存在差异，处理效应具有异质性[11]。对于这类异质性总体，现有文献大多专注于研究特定亚组的异质处理效应，而关于整个人群ATE估计的相关研究则相对较少[12-13]。此外，不同倾向性评分加权方法可能对应不同的推断目标[7]，当目标估计量与加权方法不匹配时，可能会产生偏差。

本研究通过模拟实验，探究多种常用倾向性评分加权方法在估计异质总体ATE时的有限样本性能。此外，还将这些方法应用于Lindner数据集，评估经皮冠状动脉介入治疗（percutaneous coronary intervention，PCI）联合阿昔单抗治疗对患者术后六个月内死亡率的影响。本文旨在强调依据研究问题选择合适的倾向性评分加权方法的重要性，并为处理效应异质性情况下的选择提供指导。

原理与方法

研究数据由独立同分布的随机元组$\{(X_i, Z_i, Y_i), i = 1, \cdots, N\}$构成，其中$X_i$是$p$维协变量，$Z_i$为处理变量，$Y_i$为结局变量。基于Rubin的潜在结果框架[14]，每个受试者有$Y(1)$和$Y(0)$两个潜在结果，并遵循一致性（consistency）、正值性（positivity）和条件交换性（conditional exchangeability）三个可识别性假设，以此进行因果推断。

倾向性评分$e(x_i) = \Pr(Z_i = 1|X_i = x_i)$通过整合协变量的信息，具有平衡协变量分布的特性。实际研究中PS通常是未知的，可借助逻辑回归模型$e(X_i;\ \beta) = 1/(1 + \exp(-X_i^T\beta))$进行估计。基于估计得到的PS，计算每个研究对象的平衡权重w_i，以构建一个伪总体，实现组间协变量的分布均衡[8]，进而通过下式估计ATE：

$$\hat{\Delta} = \frac{\sum_{i=1}^{N} Z_i Y_i \omega_i}{\sum_{i=1}^{N} Z_i \omega_i} - \frac{\sum_{i=1}^{N} (1 - Z_i) Y_i \omega_i}{\sum_{i=1}^{N} (1 - Z_i) \omega_i} \quad (1)$$

1. 稳定逆概率加权（stabilized inverse probability weighting，SIPW）

SIPW是在标准逆概率加权基础上发展而来的。标准逆概率加权分别为处理组和对照组分配权重$1/e_i$和$1/(1 - e_i)$，达到平衡组间协变量分布的目的，但其易受极端权重影响，导致偏差和方差较大。SIPW重新对权重进行调整使得权重之和为1，减少了极端权重的影响并提高了稳定性[6]。SIPW的目标推断总体是整个研究人群。

2. 修剪的逆概率加权（IPW with trimming，IPWT）

IPWT的基本原理是排除PS在阈值$[a, 1-a]$外的个体，此时处理组与对照组对应的权重分别为$I(a < e(x) < 1 - a)e(x)^{-1}$和$I(a < e(x) < 1 - a)(1 - e(x))^{-1}$。由于阈值$a$的设定会影响该方法的性能，本研究考虑了$a$=0.05、0.10、0.15三个阈值[15]，以探究其对处理效应的估计效果。需注意，经过修剪后的目标人群与原研究总体已不一致，因此，在估计异质性总体平均处理

* 基金项目：上海市市级科技重大专项（ZD2021CY001）；国家自然科学基金（82473724）
1. 复旦大学公共卫生学院生物统计学教研室（200032）
2. 复旦大学公共卫生学院，上海市重大传染病和生物安全研究院
#共同第一作者
△通信作者：秦国友，E-mail：gyqin@fudan.edu.cn

效应时会存在偏差。

3. 重叠加权（overlap weighting，OW）

OW[10]将权重替换为$(1-e_i, e_i)$，有效地避免了使用逆概率进行加权，克服了极端权重带来的不利影响。具体而言，PS接近0.5的个体对效应估计的贡献最大，而PS接近0或1的个体贡献最小。OW的目标人群为PS分布重叠的人群，故在估计异质性总体时，可能出现偏倚。

4. 经验似然加权（empirical likelihood weighting，ELW）

ELW旨在避免直接使用逆概率作为权重，有效克服其不稳定性。本文考虑两种ELW方法。

第一种是基于Han等[5]提出的多重稳健估计器，但在此我们只考虑一种正确指定的PS模型，得到处理组和对照组的权重分别为：

$$\hat{\omega}_i = \frac{1}{n_1}\frac{1}{1+\hat{\rho}\left(e\left(X_i;\hat{\beta}\right)-\hat{\mu}\right)} \tag{2}$$

$$\hat{\omega}_i = \frac{1}{n_2}\frac{1}{1+\hat{\eta}\left(\hat{\mu}-e\left(X_i;\hat{\beta}\right)\right)} \tag{3}$$

其中，n_1和n_2分别表示处理组和对照组的样本量，$\hat{\mu}=N^{-1}\sum_{i=1}^{N}e\left(X_i;\ \hat{\beta}\right)$，$\hat{\rho}$和$\hat{\eta}$为待估参数，该方法计算过程相对复杂。

第二种是基于Liu等[6]提出的有偏样本ELW，得到处理组和对照组的权重分别为：

$$\hat{\omega}_i = \frac{1}{n_1}\frac{1}{1+\lambda(\hat{\alpha}_1)\left(e\left(X_i;\hat{\beta}\right)-\hat{\alpha}_1\right)} \tag{4}$$

$$\hat{\omega}_i = \frac{1}{n_2}\frac{1}{1+\lambda(\hat{\alpha}_0)\left(\hat{\alpha}_0-e\left(X_i;\hat{\beta}\right)\right)} \tag{5}$$

其中，$\lambda(\hat{\alpha}_1)=n_2/(N(1-\hat{\alpha}_1))$，$\lambda(\hat{\alpha}_0)=n_1/(n_2\hat{\alpha}_0)$，$\hat{\alpha}_1$和$\hat{\alpha}_0$均为待估参数，相比于第一种方法，此方法的计算更为简便，只需求解单变量有理方程。两种ELW方法的推断目标均为整个研究人群。

5. 协变量平衡倾向性评分加权（covariate balancing propensity score weighting，CBPS）

CBPS通过优化协变量平衡来改进PS的估计，再结合逆概率加权，权重仍为$(1/e_i, 1/(1-e_i))$，其推断目标是整个研究人群，可借助R软件的CBPS包来实现具体功能。

模拟实验

本研究借助模拟生成的数据集，对上述五种倾向性评分加权方法在不同场景下的性能进行比较分析。鉴于熵平衡法是一种基于最大熵原理的权重估计方法，并且在协变量平衡方面具有较好的表现[16]，本研究亦将其纳入此次比较分析，以更全面地评估各方法在模拟场景中的性能。

1. 数据生成

参考Li等[15]文献中的模拟，对于每个研究对象i，首先从均值为μ，协方差矩阵为Σ的多元正态分布中生成协变量$V_i=(V_{1i}, V_{2i}, V_{3i}, V_{4i}, V_{5i}, V_{6i})^T$，$V_i \sim N(\mu, \Sigma)$。其中，

$$\mu=\begin{pmatrix}0\\0\\0\\0\\0\\0\end{pmatrix},\Sigma=\begin{pmatrix}1 & 0.5 & \cdots & 0.5\\0.5 & 1 & & 0.5\\\vdots & & \ddots & \vdots\\0.5 & 0.5 & \cdots & 1\end{pmatrix}$$

将前三维保持连续变量，后三维根据示性函数$I_{X<0}$转换为二元变量，从而得到协变量$\boldsymbol{X}_i=(X_{1i}, X_{2i}, X_{3i}, X_{4i}, X_{5i}, X_{6i})^T$，其期望为$(0, 0, 0, 0.5, 0.5, 0.5)^T$。我们使用logistic模型生成真实的PS：

$$PS_i = e(X_i) = (1+\exp(-(\alpha_0+\alpha_1X_{1i}+\alpha_2X_{2i}+\alpha_3X_{3i}+\alpha_4X_{4i}+\alpha_5X_{5i}+\alpha_6X_{6i})))^{-1}$$

为探究PS模型误设对上述方法的影响，进一步采用以下方式生成PS：

$$PS_i = e(X_i) = (1+\exp(-(\alpha_0+\alpha_1X_{1i}+\alpha_2X_{2i}+\alpha_3X_{3i}+\alpha_4X_{4i}^2+\alpha_5X_{5i}^2+\alpha_6X_{6i}^2)))^{-1}$$

处理变量Z_i由伯努利分布$Z_i \sim Ber(e(X_i))$产生。最后，我们使用标准差为1的正态分布生成结局变量Y_i，其均值满足：

$$E(Y_i|Z_i, X_i)=\beta_0+\beta X_i+\Delta_iZ_i$$

设置$\beta_0=0$，$\beta=(0.5, 0.5, 1.5, 0.8, 0.8, 1.0)$，$\Delta_i$表示研究对象$i$的个体处理效应。我们的目标估计参数为$\Delta=E(\Delta_i)$。

为比较各估计器性能，我们同时考虑处理效应同质和异质的情况。同质时每个研究对象$\Delta_i=0.75$；异质时设置如下六种场景：

场景1 $\Delta_i=0.75+(X_{1i}+X_{2i}+X_{3i}+X_{4i}+X_{5i}+X_{6i})\times 0.4$

场景2 $\Delta_i=0.75+(X_{1i}+X_{2i}+X_{3i}+X_{4i}+X_{5i}+X_{6i})\times 0.8$

场景3 $\Delta_i=0.75+(X_{1i}+X_{2i}+X_{3i}+X_{4i}+X_{5i}+X_{6i})\times 1.2$

场景4 $\Delta_i=0.75+(X_{1i}+X_{2i}+X_{3i}+X_{4i}+X_{5i}+X_{6i})\times 1.2-(X_{1i}X_{2i})$

场景5 $\Delta_i=0.75+(X_{1i}+X_{2i}+X_{3i}+X_{4i}+X_{5i}+X_{6i})\times 1.2-(X_{1i}^2)$

场景6 $\Delta_i=0.75+(X_{1i}+X_{2i}+X_{3i}+X_{4i}+X_{5i}+X_{6i})\times 1.2-(X_{1i}X_{2i}+X_{1i}^2)$

Δ_i表达式不同，涵盖了线性、交互项及二次项的

情形。

受Li等[15]的启发，在倾向性评分模型中采用三组参数组合 $(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5,\alpha_6)$=（0.1，0.25，0.25，−0.1，−0.25，−0.25）；（0.2，0.5，0.5，−0.2，−0.5，−0.5）；（0.45，0.6，0.6，−0.6，−0.6）分别表示良好、中等、较差的PS重叠程度，得到处理组与对照组间的PS分布如图1所示。截距 α_0 通过迭代确定，需满足期望 $E(e(X_i))$ 趋近于接受处理的概率0.4。此外，研究采用相对偏差，即绝对偏差与估计量真值之比和相对均方根误差（root mean square error，RMSE），即原始未调整估计器的均方根误差与对应各估计器的均方根误差之比）作为评价指标。对上述每种情形模拟1000个数据集，样本量设定为 N=800。

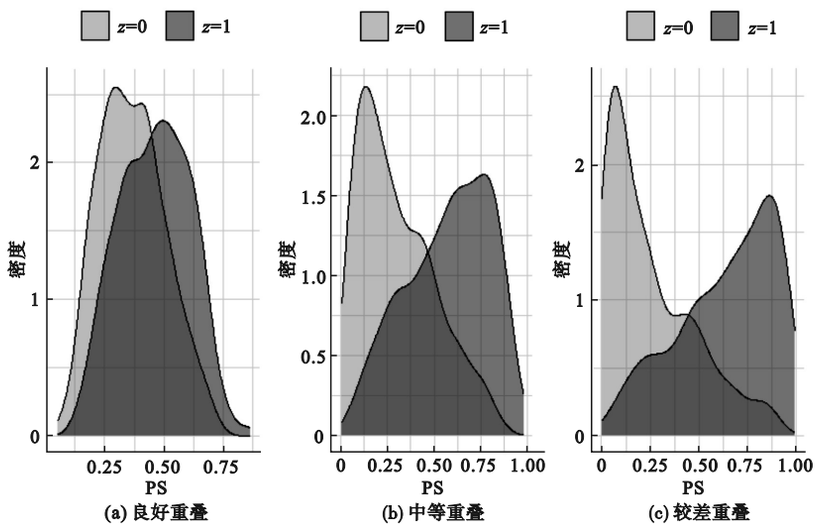


图1　接受处理概率为0.4时不同重叠程度的PS分布图（z=0为对照组，z=1为处理组）

2. 结果与分析

表1和表2分别展示了在模型正确指定时，各估计器相对于真实处理效应的相对偏差和相对RMSE，以此综合评估各方法性能，相对RMSE值越大代表性能越好。表3和表4呈现了PS模型错误指定情形下的相应结果。

在模型正确指定时，熵平衡（entropy balancing，EB）法大多数情形下均表现最优，CBPS次之。熵平衡法基于最大熵原理，通过在满足协变量矩条件下最大化权重的熵来确定权重，能够有效平衡处理组和对照组在协变量上的分布，因此展现出较好的性能。对于五种倾向性评分加权方法，推断目标为整个人群的估计器（如SIPW、ELW、CBPS）在PS分布重叠性较好且异质性较弱时表现较好，但在重叠性较差且异质性较强时开始出现偏差，相对RMSE逐渐减小。针对特定研究对象的估计器（如OW、IPWT），受异质性的影响更大，偏差随异质性的增强明显增大。这是因为在处理效应异质的人群中，Δ_{OW}（Δ_{IPWT}）与真实处理效应Δ存在偏差[15]。具体而言，CBPS估计器在场景1、2、3中表现较好，但在后三种场景中出现偏差，受Zagar等[9]的启发，原因可能是其仅平衡了协变量一阶矩，未调整由交互项 X_1X_2 及二次项引起的混杂。SIPW在重叠性较差时易产生极端权重，导致有偏估计和大方差。OW在异质性增强时偏差增大，相对RMSE减小，适用性降低。两种ELW方法中，ELW. Han相对ELW. Liu表现较好，仅次于CBPS。

在模型错误指定时，五种倾向性评分加权方法均出现有偏估计，且随着异质性增强和PS分布重叠性变差，偏差愈发明显。而熵平衡法由于通过直接设定协变量矩的平衡约束条件调整权重，能够避免因PS模型错误指定导致的协变量失衡，所以在PS模型错误指定时仍能保持较好的估计性能。

总体而言，从相对偏差和相对RMSE来看，熵平衡法不仅在模型正确指定时具有优势，在PS模型错误指定时也更为稳健。对于五种倾向性评分加权方法，CBPS更优，ELW. Han次之。

案例分析

1. 数据描述

R软件twang包中的Linder数据源自1997年辛辛那提基督医院俄亥俄心脏健康中心，包含996例接受经皮冠状动脉介入术（percutaneous coronary intervention，PCI）的患者，且被随访至少6个月。本文使用该数据比较接受PCI联合阿昔单抗治疗与只接受PCI对患者术后六个月内死亡率的影响，其中698例患者接受PCI联合阿昔单抗治疗，298例只接受PCI。此外，我们还纳入了性别、身高、有无糖尿病、有无急性心梗、有无冠状动脉支架植入、左室射血分数、PCI涉及的血管数七个协变量进行分析。关于Linder数据的详细信息见Kereiakes等[17]的文章。

表 1 不同估计器在不同数据生成模型下的相对偏差结果

估计器		数据生成模型						
		同质性	异质性					
			场景1	场景2	场景3	场景4	场景5	场景6
良好重叠	SIPW	0. 01	0. 00	0. 00	0. 00	0. 00	0. 00	0. 01
	IPWT（a=0. 05）	0. 01	0. 00	0. 00	0. 00	0. 00	0. 00	0. 01
	IPWT（a=0. 10）	0. 01	0. 01	0. 01	0. 01	0. 02	0. 03	0. 06
	IPWT（a=0. 15）	0. 01	0. 03	0. 04	0. 05	0. 10	0. 12	0. 25
	OW	0. 00	0. 05	0. 06	0. 07	0. 13	0. 17	0. 33
	ELW. Han	0. 01	0. 00	0. 00	0. 00	0. 00	0. 00	0. 01
	ELW. Liu	0. 01	0. 00	0. 00	0. 00	0. 00	0. 00	0. 01
	CBPS	0. 00	0. 00	0. 00	0. 00	0. 00	0. 00	0. 01
	EB	0. 00	0. 00	0. 00	0. 00	0. 00	0. 00	0. 01
中等重叠	SIPW	0. 00	0. 01	0. 01	0. 00	0. 02	0. 01	0. 03
	IPWT（a=0. 05）	0. 00	0. 05	0. 06	0. 06	0. 12	0. 15	0. 33
	IPWT（a=0. 10）	0. 00	0. 09	0. 12	0. 13	0. 27	0. 34	0. 70
	IPWT（a=0. 15）	0. 00	0. 10	0. 14	0. 16	0. 34	0. 43	0. 92
	OW	0. 00	0. 07	0. 10	0. 11	0. 24	0. 30	0. 65
	ELW. Han	0. 02	0. 02	0. 02	0. 01	0. 03	0. 03	0. 06
	ELW. Liu	0. 02	0. 02	0. 02	0. 02	0. 03	0. 03	0. 07
	CBPS	0. 00	0. 01	0. 01	0. 00	0. 02	0. 02	0. 05
	EB	0. 00	0. 00	0. 00	0. 00	0. 02	0. 00	0. 04
较差重叠	SIPW	0. 03	0. 04	0. 01	0. 01	0. 04	0. 03	0. 07
	IPWT（a=0. 05）	0. 00	0. 08	0. 11	0. 12	0. 28	0. 35	0. 75
	IPWT（a=0. 10）	0. 00	0. 10	0. 12	0. 14	0. 36	0. 45	1. 00
	IPWT（a=0. 15）	0. 00	0. 09	0. 12	0. 14	0. 38	0. 49	1. 11
	OW	0. 00	0. 08	0. 10	0. 12	0. 30	0. 38	0. 86
	ELW. Han	0. 06	0. 06	0. 04	0. 03	0. 08	0. 09	0. 18
	ELW. Liu	0. 07	0. 06	0. 04	0. 04	0. 08	0. 10	0. 19
	CBPS	0. 01	0. 01	0. 01	0. 01	0. 04	0. 05	0. 13
	EB	0. 01	0. 00	0. 00	0. 00	0. 06	0. 00	0. 10

表 2 不同估计器在不同数据生成模型下的相对RMSE结果

估计器		数据生成模型						
		同质性	异质性					
			场景1	场景2	场景3	场景4	场景5	场景6
良好重叠	SIPW	8. 37	9. 78	9. 82	8. 77	6. 78	6. 64	5. 08
	IPWT（a=0. 05）	8. 40	9. 78	9. 85	8. 98	6. 80	6. 68	5. 14
	IPWT（a=0. 10）	8. 69	10. 15	10. 26	9. 49	7. 36	6. 97	5. 40
	IPWT（a=0. 15）	8. 93	9. 31	8. 39	7. 18	5. 01	5. 05	3. 72
	OW	9. 43	9. 32	7. 48	6. 21	4. 47	4. 52	3. 44
	ELW. Han	8. 84	10. 96	11. 31	10. 74	8. 26	7. 91	6. 01
	ELW. Liu	8. 52	10. 17	10. 32	9. 43	7. 21	7. 01	5. 36
	CBPS	9. 24	11. 75	12. 08	11. 42	8. 82	8. 41	6. 34
	EB	9. 11	11. 94	12. 22	12. 09	10. 63	11. 23	9. 12
中等重叠	SIPW	6. 75	7. 30	6. 67	6. 66	5. 05	4. 76	3. 64
	IPWT（a=0. 05）	10. 68	10. 75	9. 74	8. 70	6. 23	6. 39	4. 59
	IPWT（a=0. 10）	12. 38	9. 73	7. 09	5. 97	3. 72	3. 91	2. 74
	IPWT（a=0. 15）	12. 89	8. 92	6. 16	5. 01	2. 97	3. 13	2. 13
	OW	14. 43	11. 76	8. 79	7. 27	4. 18	4. 43	2. 97
	ELW. Han	10. 09	11. 21	11. 21	11. 61	8. 14	7. 84	5. 70
	ELW. Liu	9. 22	9. 99	9. 72	9. 83	7. 01	6. 86	5. 03
	CBPS	12. 34	14. 79	15. 23	14. 60	9. 72	9. 98	6. 64
	EB	11. 33	13. 94	15. 85	16. 67	13. 15	15. 90	11. 52
较差重叠	SIPW	4. 35	4. 89	4. 46	3. 88	3. 30	3. 01	2. 23
	IPWT（a=0. 05）	10. 62	9. 97	8. 14	6. 90	4. 03	4. 19	2. 86
	IPWT（a=0. 10）	12. 16	10. 01	7. 71	6. 28	3. 27	3. 40	2. 22
	IPWT（a=0. 15）	12. 83	10. 35	7. 93	6. 40	3. 06	3. 16	2. 00
	OW	14. 61	12. 55	9. 57	7. 85	3. 89	4. 02	2. 58
	ELW. Han	7. 77	8. 36	8. 49	8. 84	6. 14	5. 91	4. 63
	ELW. Liu	7. 31	7. 70	7. 72	8. 01	5. 63	5. 44	4. 28
	CBPS	10. 84	13. 50	13. 22	13. 36	8. 34	7. 84	5. 57
	EB	9. 42	11. 89	14. 05	15. 91	10. 42	14. 94	10. 02

表 3　不同估计器在PS模型错误设定下的相对偏差结果

估计器		数据生成模型						
		同质性	异质性					
			场景 1	场景 2	场景 3	场景 4	场景 5	场景 6
良好重叠	SIPW	−0. 03	−0. 04	−0. 04	−0. 04	−0. 17	−0. 22	−0. 56
	IPWT（a=0. 05）	−0. 03	−0. 04	−0. 04	−0. 04	−0. 17	−0. 22	−0. 56
	IPWT（a=0. 10）	−0. 03	−0. 04	−0. 04	−0. 04	−0. 17	−0. 22	−0. 56
	IPWT（a=0. 15）	−0. 03	−0. 04	−0. 04	−0. 04	−0. 17	−0. 22	−0. 56
	OW	0. 01	0. 00	−0. 01	−0. 01	−0. 13	−0. 16	−0. 48
	ELW. Han	−0. 02	−0. 03	−0. 03	−0. 03	−0. 16	−0. 20	−0. 54
	ELW. Liu	−0. 03	−0. 03	−0. 04	−0. 04	−0. 17	−0. 22	−0. 56
	CBPS	0. 01	−0. 01	−0. 02	−0. 02	−0. 14	−0. 18	−0. 50
	EB	0. 01	0. 00	0. 00	0. 00	0. 00	−0. 00	0. 00
中等重叠	SIPW	−0. 10	−0. 10	−0. 10	−0. 10	−0. 32	−0. 40	−0. 97
	IPWT（a=0. 05）	−0. 10	−0. 10	−0. 10	−0. 10	−0. 32	−0. 40	−0. 97
	IPWT（a=0. 10）	−0. 10	−0. 10	−0. 10	−0. 10	−0. 32	−0. 40	−0. 97
	IPWT（a=0. 15）	−0. 10	−0. 10	−0. 10	−0. 10	−0. 32	−0. 40	−0. 97
	OW	0. 00	−0. 02	−0. 03	−0. 03	−0. 23	−0. 29	−0. 79
	ELW. Han	−0. 08	−0. 08	−0. 08	−0. 08	−0. 29	−0. 38	−0. 93
	ELW. Liu	−0. 09	−0. 09	−0. 10	−0. 10	−0. 31	−0. 40	−0. 96
	CBPS	0. 00	−0. 02	−0. 03	−0. 04	−0. 24	−0. 31	−0. 81
	EB	0. 00	0. 00	0. 00	0. 00	0. 01	0. 00	0. 02
较差重叠	SIPW	−0. 15	−0. 15	−0. 16	−0. 15	−0. 43	−0. 62	−1. 38
	IPWT（a=0. 05）	−0. 15	−0. 15	−0. 16	−0. 15	−0. 43	−0. 62	−1. 38
	IPWT（a=0. 10）	−0. 15	−0. 15	−0. 16	−0. 15	−0. 43	−0. 62	−1. 38
	IPWT（a=0. 15）	−0. 15	−0. 15	−0. 16	−0. 15	−0. 43	−0. 62	−1. 38
	OW	0. 00	−0. 03	−0. 05	−0. 05	−0. 29	−0. 44	−1. 09
	ELW. Han	−0. 11	−0. 12	−0. 13	−0. 13	−0. 40	−0. 58	−1. 31
	ELW. Liu	−0. 14	−0. 15	−0. 15	−0. 15	−0. 42	−0. 62	−1. 37
	CBPS	0. 00	−0. 04	−0. 05	−0. 06	−0. 30	−0. 46	−1. 12
	EB	0. 01	0. 00	0. 00	0. 00	0. 02	−0. 01	0. 03

表 4　不同估计器在PS模型错误设定下的相对RMSE结果

估计器		数据生成模型						
		同质性	异质性					
			场景 1	场景 2	场景 3	场景 4	场景 5	场景 6
良好重叠	SIPW	2. 03	2. 17	2. 12	1. 98	0. 82	0. 79	0. 74
	IPWT（a=0. 05）	2. 03	2. 17	2. 12	1. 98	0. 82	0. 79	0. 74
	IPWT（a=0. 10）	2. 03	2. 17	2. 12	1. 98	0. 82	0. 79	0. 74
	IPWT（a=0. 15）	2. 03	2. 17	2. 12	1. 98	0. 82	0. 79	0. 74
	OW	2. 20	2. 65	2. 87	2. 71	1. 02	1. 00	0. 85
	ELW. Han	2. 11	2. 37	2. 37	2. 26	0. 88	0. 84	0. 77
	ELW. Liu	2. 04	2. 19	2. 14	2. 00	0. 83	0. 79	0. 74
	CBPS	2. 19	2. 66	2. 84	2. 67	0. 97	0. 94	0. 82
	EB	2. 07	2. 51	2. 87	2. 77	2. 56	2. 46	3. 23
中等重叠	SIPW	1. 54	1. 42	1. 26	1. 16	0. 53	0. 57	0. 62
	IPWT（a=0. 05）	1. 54	1. 42	1. 26	1. 16	0. 53	0. 57	0. 62
	IPWT（a=0. 10）	1. 54	1. 42	1. 26	1. 16	0. 53	0. 57	0. 62
	IPWT（a=0. 15）	1. 54	1. 42	1. 26	1. 16	0. 53	0. 57	0. 62
	OW	2. 29	2. 74	2. 65	2. 36	0. 72	0. 77	0. 76
	ELW. Han	1. 75	1. 66	1. 49	1. 37	0. 57	0. 61	0. 65
	ELW. Liu	1. 58	1. 46	1. 29	1. 19	0. 53	0. 57	0. 63
	CBPS	2. 28	2. 69	2. 55	2. 27	0. 70	0. 74	0. 74
	EB	1. 91	2. 49	2. 64	2. 76	2. 81	2. 97	4. 18
较差重叠	SIPW	1. 32	1. 04	0. 92	0. 84	0. 42	0. 46	0. 57
	IPWT（a=0. 05）	1. 32	1. 04	0. 92	0. 84	0. 42	0. 46	0. 57
	IPWT（a=0. 10）	1. 32	1. 04	0. 92	0. 84	0. 42	0. 46	0. 57
	IPWT（a=0. 15）	1. 32	1. 04	0. 93	0. 84	0. 42	0. 46	0. 56
	OW	2. 44	2. 65	2. 36	1. 99	0. 62	0. 64	0. 71
	ELW. Han	1. 57	1. 25	1. 10	0. 99	0. 46	0. 49	0. 59
	ELW. Liu	1. 37	1. 08	0. 96	0. 87	0. 43	0. 46	0. 57
	CBPS	2. 42	2. 61	2. 29	1. 93	0. 60	0. 62	0. 69
	EB	1. 86	2. 47	2. 72	2. 72	2. 64	3. 41	5. 30

2. 异质性分析与结果

我们分别基于有无急性心梗和有无糖尿病两个变量对总体进行分组，计算亚组间ATE的差值衡量该总体的异质性程度。以有无急性心梗为例，0为无急性心梗，1为有急性心梗，基于此变量的异质性度量dATE=ATE（有急性心梗亚组）−ATE（无急性心梗亚组），结果见表5，为方便比较，所有结果均乘以100。此外，我们借助bootstrap法重抽样1000次，对相应的标准误进行估计。由表5亦知，该人群存在较为明显的处理效应异质性。

表5 PCI联合阿昔单抗治疗对术后六个月内死亡率的效应分析与比较

估计器	总体处理效应		亚组间处理效应比较：dATE			
	ATE	SE	有无急性心梗		有无糖尿病	
			dATE	SE	dATE	SE
Unadjusted	3. 46	0. 018	4. 49	0. 582	3. 54	0. 136
SIPW	6. 61	0. 086	19. 48	3. 989	10. 31	0. 695
IPWT（a=0. 05）	4. 44	0. 055	−3. 23	1. 089	9. 91	0. 419
IPWT（a=0. 10）	3. 21	0. 025	0. 71	0. 579	−1. 35	0. 286
IPWT（a=0. 15）	2. 63	0. 018	5. 84	0. 826	−0. 31	0. 223
OW	4. 01	0. 020	6. 96	0. 738	3. 63	0. 178
ELW. Han	6. 27	0. 060	25. 72	4. 130	7. 37	0. 310
ELW. Liu	6. 47	0. 069	22. 65	3. 787	8. 98	0. 384
CBPS	6. 18	0. 048	17. 37	2. 691	6. 89	0. 293
EB	5. 63	0. 054	10. 05	1. 127	6. 89	0. 528

随后，我们采用上述方法进行ATE的估计，结果见表5相应部分。理论上，推断目标并非整个人群的估计器（如OW、IPWT等），对该异质性总体ATE的估计会存在偏差。结果显示，SIPW、ELW. Han（Liu）和CBPS的估计值十分接近，而OW、IPWT的估计有一定差异，这源于推断目标的不同，也验证了我们的结论：对于异质性总体，推断目标并非整个人群的估计器，不适合用于估计整个人群的ATE。

讨　论

在观察性研究中，调整混杂因素对于准确估计ATE至关重要，倾向性评分加权方法是常用的手段之一。本研究聚焦于评估不同倾向性评分加权方法在调整混杂因素时的性能表现，具体考察在多种异质性情形以及PS重叠程度下这些方法的效果。此前，已有一些研究对部分倾向性评分加权方法进行了比较分析。然而，这些研究多局限于同质总体的情境[5-6,15,18]。虽也有文献涉及异质总体，但所设置的异质性场景较为简单，且估计目标并非整个人群的ATE[19-20]。鉴于此，本研究设计了六种不同的异质性场景进行模拟实验，深入地探究了各加权方法在估计异质总体ATE时的性能优劣。

结果表明，CBPS在各种异质性场景以及PS重叠程度下，表现均优于其他估计器，是实现估计目标的最佳选择。具体而言，当异质性较弱且重叠良好时，各估计器均能提供较为理想的估计结果。然而，随着异质性增强、重叠性变差，估计器性能开始分化。CBPS估计器相对而言表现更好，IPWT和OW估计器由于推断目标并非整个人群，在这种复杂情境下，偏差会明显增大。SIPW在重叠性较差时，易受到极端权重的影响，进而导致估计有偏且方差较大。ELW. Liu在文献中被证明最大化权重比$k=(\max_{1\leqslant i\leqslant n_1} w_i)/(\min_{1\leqslant i\leqslant n_1} w_i)$（对照组同理）存在上界，但这个上界总是非常大，在本研究模拟中未达到，致使其良好性能未能体现。而ELW. Han由于估计目标为整个人群且避免在权重中使用逆概率，在模拟中表现较优，仅次于CBPS。

总体而言，本研究借助模拟实验深入评估了多种倾向性评分加权方法在异质性情景下的表现，明确了不同方法在估计异质性总体ATE时的特性，为在观察性研究中选择合适的倾向性评分加权方法提供了参考。

参 考 文 献

[1] Hernán MA, Robins JM. Causal Inference: What If[M]. Boca Raton: Chapman & Hall/CRC, 2020.

[2] 孙婷，秦国友，武振宇，等. 不同混杂结构条件下各倾向性评分方法的模拟比较研究[J]. 中国卫生统计，2017，34(3)：415-420.

[3] Crump RK, Hotz VJ, Imbens GW, et al. Dealing with limited overlap in estimation of average treatment effects[J]. Biometrika, 2009, 96(1): 187-199.

[4] Imai K, Ratkovic M. Covariate balancing propensity score[J]. Journal of the Royal Statistical Society Series B, Statistical methodology, 2014, 76(1): 243-263.

[5] Han P, Wang L. Estimation with missing data: beyond double robustness[J]. Biometrika, 2013, 100(2): 417-430.

[6] Liu Y, Fan Y. Biased-sample empirical likelihood weighting for missing data problems: an alternative to inverse probability weighting[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2023, 85(1): 67-83.

[7] Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the

propensity score: a primer for practitioners[J]. BMJ, 2019, 367: l5657-l5657.

[8] 涂博祥,秦婴逸,吴骋,等. 倾向性评分加权方法介绍及R软件实现[J]. 中国循证医学杂志,2022,22(3):365-372.

[9] Zagar AJ, Kadziola Z, Lipkovich I, et al. Evaluating different strategies for estimating treatment effects in observational studies[J]. Journal of Biopharmaceutical Statistics, 2017, 27(3): 535-553.

[10] Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting[J]. Journal of the American Statistical Association, 2018, 113(521): 390-400.

[11] Brooks JM, Chapman CG, Schroeder MC. Understanding treatment effect estimates when treatment effects are heterogeneous for more than one outcome[J]. Applied Health Economics and Health Policy, 2018, 16(3): 381-393.

[12] Robertson SE, Leith A, Schmid CH, et al. Assessing heterogeneity of treatment effects in observational studies[J]. American Journal of Epidemiology, 2021, 190(6): 1088-1100.

[13] Dahabreh IJ, Hayward R, Kent DM. Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence[J]. International Journal of Epidemiology, 2016, 45(6): 2184-2193.

[14] Rubin DB. Estimating causal effects from large data sets using propensity scores[J]. Annals of Internal Medicine, 1997, 127(8 Part 2): 757-763.

[15] Li F, Thomas LE. Addressing extreme propensity scores via the overlap weights[J]. American Journal of Epidemiology, 2019, 188(1): 250-257.

[16] Hainmueller J. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies[J]. Political Analysis, 2012, 20(1): 25-46.

[17] Kereiakes DJ, Obenchain RL, Barber BL, et al. Abciximab provides cost-effective survival advantage in high-volume interventional practice[J]. American Heart Journal, 2000, 140(4): 603-610.

[18] Han P. A further study of the multiply robust estimator in missing data analysis[J]. Journal of Statistical Planning and Inference, 2014, 148: 101-110.

[19] Zhou YJ, Matsouaka RA, Thomas L. Propensity score weighting under limited overlap and model misspecification[J]. Statistical Methods in Medical Research, 2020, 29(12): 3721-3756.

[20] Garrido MM, Lum J, Pizer SD. Vector-based kernel weighting: A simple estimator for improving precision and bias of average treatment effects in multiple treatment settings[J]. Statistics in Medicine, 2021, 40(5): 1204-1223.

（责任编辑：林美艺）

（上接第713页）

[23] Yang S, Lin S, Li N, et al. Burden, trends, and risk factors of esophageal cancer in China from 1990 to 2017: an up-to-date overview and comparison with those in Japan and South Korea[J]. Journal of Hematology & Oncology, 2020, 13(1): 146.

[24] Shigaki H, Baba Y, Watanabe M, et al. LINE-1 hypomethylation in noncancerous esophageal mucosae is associated with smoking history[J]. Annals of Surgical Oncology, 2012, 19(13): 4238-4243.

[25] Jin EH, Han K, Shin CM, et al. Sex and Tumor-Site Differences in the Association of Alcohol Intake With the Risk of Early-Onset Colorectal Cancer[J]. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 2023, 41(22): 3816-3825.

[26] 国务院第七次全国人口普查领导小组办公室. 中国人口普查年鉴2020[M]. 北京:中国统计出版社, 2022.

[27] 地力亚尔·吾斯曼江. 1990—2019年归因于吸烟、饮酒、高BMI的中国食管癌死亡率趋势分析[D]. 乌鲁木齐: 新疆医科大学, 2023.

[28] 陈思念, 曾智, 朱嘉雯, 等. 中国归因于饮酒的食管癌疾病负担分析与预测:基于1990—2019年中国疾病负担的年龄-时期-队列研究[J]. 肿瘤, 2024, 44(2): 158-168.

（责任编辑：林美艺）