

《药物流行病学研究方法学指南（第2版）》及其系列解读（15）：倾向评分分析的要点与案例



胡源晖^{1, 2, 3, 4}, 刁莎^{1, 2, 3}, 伯贞艳^{1, 2, 3}, 邹 锐^{1, 2, 3}, 曾力楠^{1, 2, 3, 5}, 李海龙^{1, 2, 3},
张伶俐^{1, 2, 3, 5, 6}

1. 四川大学华西第二医院药学部/循证药学中心（成都 610041）
2. 儿童药物四川省重点实验室（成都 610041）
3. 出生缺陷与相关妇儿疾病教育部重点实验室（成都 610041）
4. 四川大学华西药学院（成都 610041）
5. 四川大学华西生物医学大数据研究院（成都 610041）
6. 四川大学华西医院中国循证医学中心（成都 610041）

【摘要】倾向评分在药物流行病学观察性研究中被广泛用于控制混杂偏倚，以提高因果效应估计的可靠性。《药物流行病学研究方法学指南（第2版）》概述了倾向评分的基本概念及使用场景。本文在该指南基础上，围绕“协变量筛选-倾向评分估计-基本假设检验-倾向评分应用（匹配、分层、回归、加权）-协变量平衡性检验-因果效应估计-敏感性分析-规范报告”等关键步骤，通过案例阐述倾向评分分析的核心要点，旨在为临床与流行病学研究者在真实世界药物效果与安全性研究中规范选择、实施与报告倾向评分分析提供参考。

【关键词】药物流行病学；方法学；指南；倾向评分

【中图分类号】R181.3+5 **【文献标识码】**A

Guide on Methodological Standards in Pharmacoepidemiology (2nd edition) and their series interpretation (15): key points and examples of propensity score analysis

HU Yuanhui^{1,2,3,4}, DIAO Sha^{1,2,3}, BO Zhenyan^{1,2,3}, ZOU Kun^{1,2,3}, ZENG Linan^{1,2,3,5}, LI Hailong^{1,2,3},
ZHANG Lingli^{1,2,3,5,6}

1. Department of Pharmacy/Evidence-Based Pharmacy Center, West China Second University Hospital, Sichuan University, Chengdu 610041, China
2. Children's Medicine Key Laboratory of Sichuan Province, Chengdu 610041, China
3. Key Laboratory of Birth Defects and Related Diseases of Women and Children, Sichuan University, Ministry of Education, Chengdu 610041, China
4. West China School of Pharmacy, Sichuan University, Chengdu 610041, China
5. West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu 610041, China
6. Chinese Evidence-based Medicine Center, West China Hospital, Sichuan University, Chengdu 610041, China

DOI: 10.12173/j.issn.1005-0698.202603031

基金项目：国家自然科学基金面上项目（72574158）；四川省重点研发项目（2024YFFK0082）

通信作者：张伶俐，博士，教授，主任药师，博士研究生导师，Email: zhanglingli@scu.edu.cn

李海龙，博士，副研究员，Email: lhl_stat@scu.edu.cn

Corresponding authors: ZHANG Lingli, Email: zhanglingli@scu.edu.cn; LI Hailong, Email: lhl_stat@scu.edu.cn

【Abstract】 Propensity scores are widely used in observational pharmacoepidemiology studies to control confounding bias and improve the reliability of causal effect estimation. *The Guide on Methodological Standards in Pharmacoepidemiology (2nd edition)* summarizes the basic concepts and application scenarios of propensity scores. Building upon this guideline, this paper elaborates on the key practical elements of propensity score analysis through examples, following a standardized workflow encompassing covariate selection, propensity score estimation, assessment of basic assumptions, propensity score application (matching, stratification, regression, or weighting), covariate balance test, causal effect estimation, sensitivity analysis, and standardized reporting. The aim is to provide a practical reference for clinical and epidemiological researchers in conducting standardized selection, implementation, and reporting of propensity score analysis in real-world studies on drug efficacy and safety.

【Keywords】 Pharmacoepidemiology; Methodology; Guidelines; Propensity score

药物流行病学是运用流行病学原理与方法，研究人群中药品的利用及其效应的应用科学^[1]，其研究结论的可靠性高度依赖于方法学设计的科学性与严谨性。为促进研究方法的规范与统一，提升我国药物流行病学研究质量，2024年中国药学会启动了《药物流行病学研究方法学指南（第2版）》（以下简称“指南第2版”）的制订/修订工作^[2]，并于2025年10月正式发布^[3]。指南第2版强调了倾向评分（propensity score, PS）在药物流行病学研究中的重要作用。本文主要针对指南第2版中PS分析的要点进行解读，旨在为相关研究者提供系统的方法学参考与实践指导。

1 PS的定义与基本假设

PS是指在已知所有已测量的、可能影响研究结局的协变量条件（如年龄、病情严重程度、合并用药等）下，某个个体接受目标暴露（如特定药物治疗）的预测概率^[4]。其核心作用是，通过估计每个个体接受暴露的可能性，在后续分析（如加权或匹配）中构建一个近似随机化分配的“伪研究总体”，从而减少因适应证混杂等因素导致的系统性偏倚，以便更可靠地评估暴露与结局之间的因果关系^[5-6]。

PS的有效应用需满足因果推断中的两个基本假设：条件可交换性假设与正性假设。条件可交换性假设要求在考虑所有已测量的协变量后，暴露组和对照组之间具有可比性^[7]。换言之，通过这些协变量对人群进行充分调整后，个体是否接受暴露可以近似看作是随机分配的，如同随机对照试验^[8]。正性假设要求每一个个体，无论其协

变量特征如何，都有机会（非零概率）接受任何一种暴露状态^[8]。换句话说，在每一组协变量取值下，既要有接受暴露的个体，也要有未接受暴露的个体。如果某个协变量组合下的个体全部属于暴露组（或全部属于对照组），那么就没有可供参照的比较对象，也就无法可靠估计因果效应。

2 PS分析

PS分析通常遵循一套标准化流程，主要包括协变量筛选、PS估计、基本假设检验、PS应用、组间协变量平衡性检验、因果效应估计、敏感性分析及规范报告等步骤。

2.1 协变量筛选

PS分析的首要步骤是基于既往文献、专业知识及因果推断框架（如有向无环图），识别并纳入可能构成混杂的协变量，即同时影响暴露和结局的变量；此外，也可纳入仅与结局相关的预测变量以提高估计效率^[9-10]。当缺乏足够的先验知识构建完整因果有向无环图时，也可辅助使用数据驱动方法进行协变量筛选。常见的数据驱动方法包括基于显著性水平（如 P 值）的筛选、逐步回归法、最小绝对收缩和选择算子（least absolute shrinkage and selection operator, LASSO）回归等。具体方法本文不展开介绍，相关方法学细节可查阅相应书籍及文献。

在协变量筛选过程中，必须审慎排除几类特定变量：首先，应排除仅影响目标暴露的工具变量（如医院用药政策），因其不构成混杂，控制此类变量会放大未测量混杂带来的偏倚^[11]；其次，应排除作为目标暴露与结局共同结果的碰撞

变量（如研究吸烟与肺癌关系中的生活质量），控制此类变量将激活虚假的非因果关联路径，引入新的混杂偏倚^[12]；最后，位于暴露至结局因果路径上的中介变量（如研究吸烟与肺癌关系中的慢性支气管炎）也应被剔除，因其代表了暴露作用的一部分，控制它们将掩盖真实的总体因果效应^[13]。

筛选协变量时还应关注协变量与目标暴露的时间先后顺序。所有协变量的测量时点必须明确早于或至少不晚于目标暴露的发生时点，以确保它们在逻辑上可能成为真实的混杂因素，而非目标暴露的结果或中介变量。在探讨暴露随时间变化对结局的影响时，建议采用目标暴露发生前的最近一次测量值作为对应时间点的时变协变量，

并可同时控制既往目标暴露水平，以更有效地识别并调整时变混杂^[14]。

图 1 为用于探讨“糖皮质激素使用”对“肺炎死亡”影响的因果有向无环图^[15]，该图未考虑时依协变量，各节点含义及作用包括：①混杂因素（如年龄、性别）因同时影响治疗决策与结局，需纳入 PS 模型校正；②竞争暴露（如利尿剂）通过独立路径影响结局（如减轻肺水肿），与激素治疗无直接关联，纳入模型可提升拟合精度；③中介变量（如血压、肌酐）位于因果路径中，不应纳入模型；④碰撞变量（如住院天数）由暴露与结局共同影响，校正会引入碰撞偏倚；⑤工具变量（如医生处方偏好）仅影响治疗概率，与结局无直接因果关联，亦不应纳入模型。

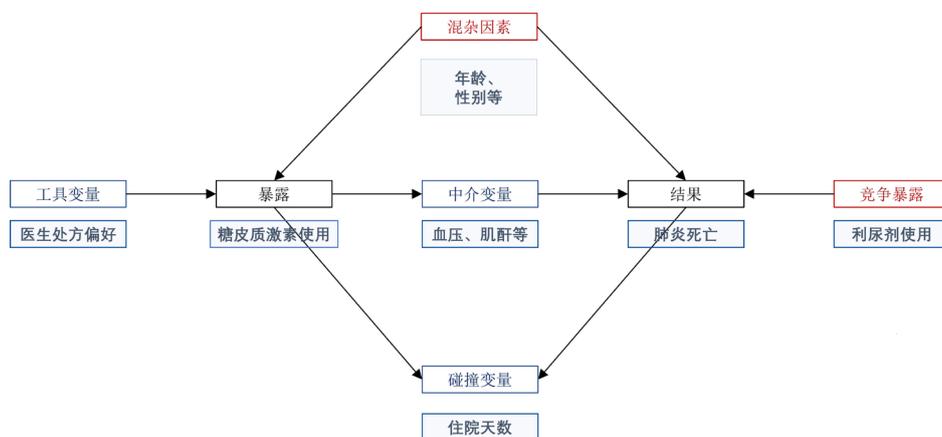


图1 因果有向无环图

Figure 1. Causal directed acyclic graph

2.2 PS估计

在筛选出合适的协变量后，可使用多种方法估计个体的 PS。最常用的 PS 估计方法为逻辑回归模型，其他估计方法包括神经网络、支持向量机、决策树等^[16]。可使用边际结构模型^[17]等方法估计时变协变量的 PS。这些方法的核心是基于所有协变量预测个体被分配到暴露组的概率，该概率即为 PS，其取值介于 0 和 1 之间。后续将基于 PS 进行匹配、加权或分层等操作，以平衡组间协变量分布并估计因果效应。

2.3 PS基本假设的检验

PS 的有效应用依赖于条件可交换性假设与正性假设的满足。在完成 PS 估计后，需对这 2 项基本假设进行检验。

条件可交换性假设要求所有重要的混杂因素均已被测量并纳入分析^[8]。然而，观察性研究中

可能存在未测量的混杂变量，故该假设本身难以通过统计检验直接验证。因此实际应用中主要通过通过对已测量协变量进行平衡性评估，以为其成立提供间接证据，具体检验方式有 2 种：一是将样本按 PS 等分为若干区间（如五分位），分别计算各区间内暴露组与对照组在各协变量上的绝对标准化差值（absolute standardized difference, ASD），通常以 $ASD < 0.10$ 作为平衡良好的标准^[18]，亦有研究^[19]建议采用 < 0.25 作为宽松参考；二是在应用 PS 分析方法（如匹配或加权）后，对暴露组与对照组间的协变量平衡性进行检验（具体方法将在“2.5”节详细介绍），当发现协变量存在不平衡时，可通过变量变换、重新分类或纳入交互项等方式尝试解决^[20]。

正性假设通过比较暴露组与对照组之间 PS 的分布重叠情况来评估，可通过绘制直方图、核

密度图等进行可视化观察^[20-21]。以核密度图为例(图2),2组PS核密度曲线的重叠情况可归纳为4种情形:图2-A为完全不重叠,提示组间协变量不可比,严重违反正性假设;图2-B和图2-C为部分重叠,其中图2-C为子集重叠,即一组PS分布完全包含于另一组范围内;图2-D为良好重叠,此为开展因果推断的理想情形。除重叠外,暴露组和对照组之间PS的分布应相似,可通过比较各PS等分区间(如五分位)内2组PS的ASD进行评估^[22]。若出现大范围分布不重叠或多数区间内ASD超出可接受范围,则可能违反正性假设。此时可考虑扩大样本、优化协变量选择或重新界定研究人群以尝试解决正性假设缺失的问题^[23]。

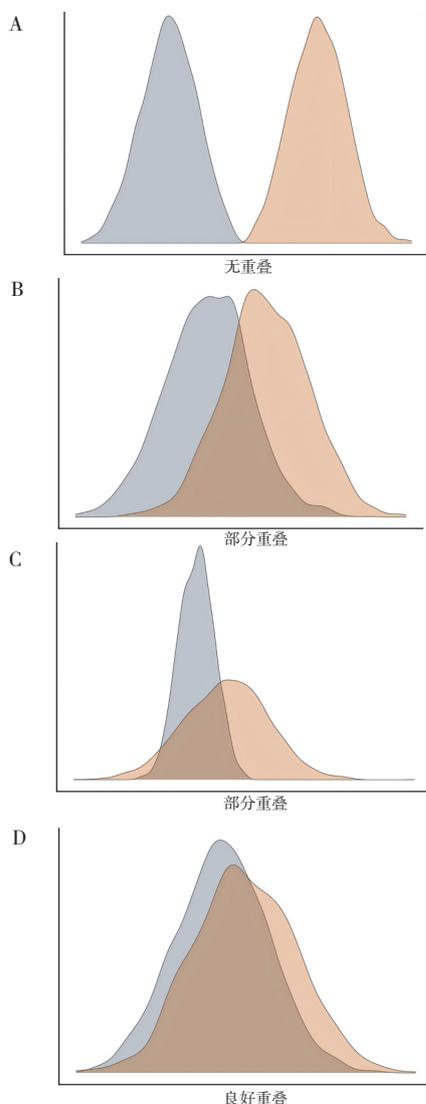


图2 核密度图示例^[5]

Figure 2. Examples of kernel density plots^[5]

注: A. 无重叠; B/C. 部分重叠; D. 良好重叠; 黄色表示暴露组, 蓝色表示对照组; 横轴为PS值, 纵轴为分布密度。

2.4 PS应用

在完成PS基本假设的验证后,下一步可通过匹配、分层、回归或加权等常用方法,以实现组间协变量均衡可比。

在PS应用前,需依据研究目的选择合适的因果效应估计量^[24](表1),包括全样本的平均处理效应(average treatment effect, ATE)、暴露组的平均处理效应(average treatment effect in the treated, ATT)、对照组的平均处理效应(average treatment effect in the untreated, ATU)以及重叠人群的平均处理效应(average treatment effect in the overlap, ATO)。ATE表示全样本所有个体接受与不接受暴露时预期结果的平均差异,用于评估暴露对整体人群的影响,如评估某新型降压药在广泛高血压人群中的总体疗效与推广价值。ATT表示暴露组所有个体接受与不接受暴露时预期结果的平均差异,用于评估暴露在实际接受暴露人群中的效应,如明确某降压药对当前已使用该药物高血压患者的实际获益。ATU表示非暴露组所有个体接受与不接受暴露时预期结果的平均差异,用于评估暴露在未接受暴露人群中的效应,如预测某降压药若推广至目前未使用该药的高血压患者群体中可能带来的额外获益。ATO表示在临床决策不确定性较高(即PS重叠区域)的人群中,所有个体接受与不接受暴露时预期结果的平均差异,用于评估暴露在PS相近,即最可能接受或不接受暴露的亚群中的效应,如评估某降压药在临床特征介于明确用药与明确不用药之间患者中的疗效与风险权衡。

2.4.1 匹配

匹配是指将暴露组中的个体与对照组中的个体配对,其目的是构建一个协变量分布均衡、组间可比的分析样本,从而实现了对因果效应的有效估计^[25](图3)。常用的匹配方法包括最近邻匹配、最优匹配、全匹配和精确匹配。

最近邻匹配为每个暴露组个体寻找PS最为接近的对照个体进行配对^[26]。与最近邻匹配不同,最优匹配通过最小化所有配对间的PS差异总和,从而实现样本整体上的协变量平衡^[27]。最优匹配与最近邻匹配在组间协变量平衡方面表现相似,但后者在减少因果效应估计值偏差和均方误差方面可能表现更优^[28]。全匹配是一种更灵活的匹配框架,旨在最大程度地利用样本数

表1 因果效应估计量及其估计方法总结^[24]

Table 1. Summary of causal effect estimators and its estimation methods^[24]

因果效应估计量	目标人群	适用研究问题示例	主要适用匹配方法	主要适用加权方法
ATT	暴露组	是否应该对接受某药物治疗的患者继续该药物治疗?	所有匹配方法、分层	标准化死亡率比加权、匹配加权
ATU	对照组	是否应该对未接受某药物治疗的患者实施该药物治疗?	同ATT	同ATT
ATE	全人群	是否应该对所有符合条件的患者实施该药物治疗?	全匹配、分层	逆概率加权、匹配加权
ATO	PS重叠人群	是否应该对处于临床均衡状态的一般风险患者实施该药物治疗?	-	重叠加权



图3 PS匹配图示

Figure 3. Diagram of PS matching

注：黄色小人为暴露组，蓝色小人为对照组。

据。它将样本划分为多个“匹配集”，每个匹配集可包含一个暴露个体与一个或多个对照个体，或反之^[29-30]。其能最大限度地保留可用样本，并通过最小化匹配集内暴露组与对照组个体之间的PS差异，以及对匹配集内个体赋予权重，最终实现组间协变量的均衡可比。精确匹配要求在所有匹配的协变量中，暴露组和对照组的个体完全相同，即每个暴露组个体都找到一个在所有匹配变量中数值完全一致的对照组个体^[31]。这种方法确保了匹配后2组协变量的完全平衡，避免了因协变量差异带来的偏倚，但可能导致样本量大幅减少。

所有匹配方法均可用于估计 ATT，因其为每个暴露组个体匹配相似的对照组个体；而全匹配因可通过权重灵活适配总体分布，还可用于估计 ATE。可结合研究目标、样本特征以及数据保留需求等选择合适的匹配方法。在选定匹配方法后，需设定具体的匹配参数，这主要包括匹配方式（1:1 匹配或 1:n 匹配）、匹配距离（卡钳值的设定，定义为允许匹配的最大距离，设定值推荐设定为 PS logit 转换值标准差的 0.2~0.25 倍^[26]）以及抽样方法（放回抽样或不放回抽样）等^[5]。

匹配的优势在于能够提供良好的协变量平衡，且其分析结果直观可靠。但由于未匹配患者

会被排除在分析之外，因此匹配可能导致系统性偏倚增加和统计效能降低^[32]。

Ramsey 等^[33]利用 TriNetX 数据库评估胰高血糖素样肽-1 受体激动剂对 2 型糖尿病患者眼部并发症的影响。研究者采用最近邻匹配以控制 2 组基线差异，具体参数设置为：匹配比例为 1:1，设置卡钳值为 PS logit 转换值标准差的 0.1 倍，抽样方法为不放回抽样。匹配后以 ASD < 0.1 评估组间协变量平衡情况，随后使用 Cox 比例风险模型估计胰高血糖素样肽-1 受体激动剂与各结局事件发生风险之间的关联。

2.4.2 分层

分层是指根据 PS 将研究对象划分为若干层，使得每一层内的暴露组和对照组的协变量分布均衡，进而在每个层内分别比较暴露组与对照组的结局差异，以更准确地估计因果效应，可用于估计 ATE 或 ATT^[34]（图 4）。通常将样本等分为 5 层^[35]，但在某些情况下，尤其是暴露较少时，使用更细的分层（如 50 或 100 层）可以提高估计的精确度和减少偏差^[36]。

与其他 PS 方法（如匹配和加权）相比，分层操作简单且易于解释，但在事件较少时可能表现较差^[37]。

Elze 等^[37]评估慢性心力衰竭患者 β 受体阻滞剂使用与全因死亡风险之间的关联，首先基于年

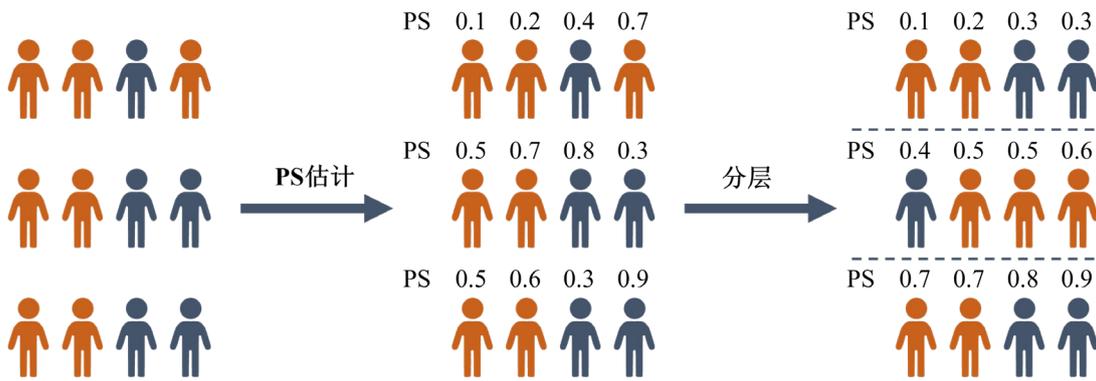


图4 PS分层图示

Figure 4. Diagram of PS stratification

注：黄色小人为暴露组，蓝色小人为对照组。

龄、性别、吸烟、糖尿病等 18 项基线协变量估计 PS；随后按 PS 大小将总体样本划分为等人数的 5 层和 10 层，并在每一层内分别估计暴露组与对照组的结局差异；最后对各层结果进行加权汇总得到总体效应估计。

2.4.3 回归

回归是指将 PS 作为协变量纳入回归模型，调整混杂因素对因果估计的影响，从而减少偏倚^[34]。应注意，相较于匹配或加权，单纯将 PS 作为协变量纳入回归模型可能对模型设定更为敏感^[38]，易受变量形式、交互效应与模型假设等影响。因此在药物流行病学研究中通常优先采用其他方法。

2.4.4 加权

加权是指利用 PS 对每个个体赋予相应的权重，通过构建一个协变量分布均衡的伪总体以模拟随机化试验的条件，可用于估计所有因果效应估计量^[39]。常见的加权方法包括逆概率处理加权(inverse probability of treatment weighting, IPTW)、标准化死亡率比加权 (standardized mortality ratio weights,

SMRW)、匹配加权 (matching weight, MW) 以及重叠加权 (overlap weight, OW) (图 5)。

IPTW 按患者实际接受暴露概率的倒数进行加权，即暴露组个体的权重为 $1/PS$ ，对照组个体的权重为 $1/(1-PS)$ ，主要用于估计 ATE。由于权重直接由 PS 计算得出，当 PS 接近 0 或 1 时，常会产生极端权重。为提高估计稳定性，实践中常采用稳定权重，即在上述权重分子中引入总体暴露概率^[40]：暴露组权重为 P_t (样本中暴露组所占比例) / PS，对照组权重为 $(1-P_t) / (1-PS)$ 。为进一步处理极端权重并防止方差膨胀，通常还需要对权重进行截断。常见的做法是根据权重的分布 (如第 1 与第 99 百分位数) 预设权重的上下限，对超出该范围的权重进行截断^[41]。通过设定更严格的截断阈值可进一步降低方差膨胀，但会引入额外偏倚^[8]。

SMRW 是将暴露组患者的权重设为 1，对照组患者按其接受暴露概率与未接受暴露概率的比值 $[PS / (1-PS)]$ 进行加权，使其代表暴露组人

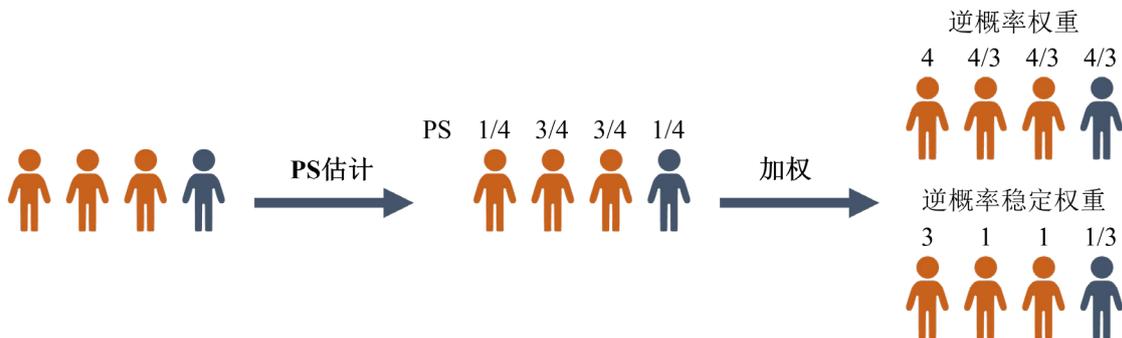


图5 PS加权图示 (以IPTW为例)

Figure 5. Diagram of PS weighting (taking IPTW as an example)

注：黄色小人为暴露组，蓝色小人为对照组。

群, 从而估计 ATT^[42]。与 IPTW 类似, 由于直接使用 PS 计算权重, SMRW 也可能产生极端权重。若观察到较大的权重, 可考虑进行权重截断。在多组比较中, 研究者可通过将目标暴露组权重设为 1, 并以此组 PS 与其他组实际接受暴露的 PS 之比为其他组加权, 以实现特定目标暴露组的 ATT 估计。

MW 基于患者接受暴露与不接受暴露的 PS 中较小者与实际接受暴露的 PS 的比值对患者进行加权^[43], 即暴露组个体的权重为 $\min(PS, 1-PS)/PS$, 对照组个体的权重为 $\min(PS, 1-PS)/(1-PS)$ 。由于其设计使得权重被限制在 0 到 1 之间, 因此不可能出现极端权重, 从而无需进行权重截断。当各组样本量相近时, 其因果效应估计量接近于全人群中的平均处理效应; 当各组样本量不等时, 其因果效应估计量接近于观测数较少那一组 (通常是暴露组) 人群中的平均处理效应。其多组扩展公式类似, 分子为所有 PS 中的最小值, 分母为实际接受暴露的 PS。

OW 基于患者接受相反暴露的 PS 对其进行加权^[44], 即暴露组个体的权重为 $1-PS$, 对照组个体的权重为 PS 。重叠权重的目标人群是两组中 PS 相近的重叠人群, 因此其主要用于估计 ATO。与 MW 类似, 由于权重被限制在 0 到 1 之间, 不可能出现极端权重, 因此无需进行权重截断。其多组扩展形式称为“广义重叠权重”, 本质上是基于广义 PS 构造的一类平衡权重, 用于突出在所有暴露组中均具有较高分配可能性的重叠人群^[45]。

相较于匹配, 加权保留了分析中的绝大多数样本, 因此在估计因果效应时可以提供更高的精确度。实际操作中应结合研究目的选择适配的权重类型。

Xu 等^[46] 基于美国国家健康与营养检查调查 (National Health and Nutrition Examination Survey, NHANES) 数据, 在成人人群中评估了苯二氮草类药物 (单用或与阿片类合用) 与全因死亡风险之间的关联。为尽量消除不同用药人群的基线差异, 研究者首先使用包含社会人口学因素、合并症、处方信息等的协变量估计 PS; 随后采用 SMRW 构建加权伪总体: 将对照组权重设为 1, 暴露组按 $(1-PS)/PS$ 加权; 加权后以 $ASD < 0.1$ 检验协变量平衡, 并在加权样本上使用 Cox 比例

风险模型进行加权后风险比的估计。

2.5 组间协变量平衡性检验

PS 应用后, 需评估暴露组与对照组间协变量的平衡情况。可采用 ASD 进行定量评估, 通常认为 $ASD < 0.10$ 表示协变量平衡良好。此外, 也可借助图示法 (如 Q-Q 图、密度曲线重叠图、箱线图等) 定性评估组间协变量的分布差异^[18]。

2.6 因果效应估计

在证实组间协变量达到充分平衡后, 即可进行因果效应估计。

应根据结局变量的类型选择相应的模型框架: 对于连续型或二分类结局, 通常采用广义线性模型; 对于时间至事件结局, 需采用 Cox 比例风险模型。在选定相应的结局模型框架后, 需根据不同 PS 方法的特点, 实施相应的分析操作: 若采用分层, 需在各 PS 层内分别估计效应, 再以层内样本量或其占比为权重合并得到总体效应^[47]; 若采用匹配, 可在匹配后的样本中直接构建模型进行估计, 并建议在模型中纳入匹配对以提高精度^[48]; 若采用加权, 则需基于 PS 的权重进行加权回归^[47]。

在分析涉及多次测量和时变暴露的纵向数据时, 可采用边际结构模型等专门框架, 通过时变逆概率加权来处理时变混杂^[17]。

无论采用何种方法, 均建议在结局模型中纳入在平衡诊断中仍存在残余不平衡的协变量, 以进一步控制混杂^[49]。

2.7 敏感性分析

敏感性分析是 PS 分析的重要组成部分, 其核心目的为评估未观测混杂因素对结果的影响, 确保结果的稳健性。可采用的方法包括估计 E 值 (E-value)^[50] 或构建基于 PS 的敏感性函数^[51]等。

此外, PS 模型与结局模型的设定涉及一系列方法学选择, 如协变量纳入策略、模型形式选择等。建议采用敏感性分析 (如更换模型设定、调整协变量纳入策略或采用不同的 PS 估计方法) 来系统评估不同方法学选择对因果效应估计结果的影响。

2.8 PS 报告

在使用 PS 方法的研究中, 翔实的报告至关重要。除遵循观察性研究的报告规范 (如 STROBE 声明) 外, 还需补充以下内容^[52] (表 2):

在方法部分, 首先应明确界定因果效应估计

表2 PS报告建议

Table 2. Instructions for reporting PS analysis

研究部分	报告建议
方法部分	(1) 明确界定所关注的因果效应估计量 (2) 详细描述PS估计模型，具体说明模型类型、纳入的协变量以及是否包含非线性项或交互项 (3) 详细描述选择协变量的方法 (4) 详细描述检验PS关键假设所采用的方法 (5) 若采用PS匹配，需清晰阐述匹配具体方法，包括匹配算法、卡钳值、匹配比例等 (6) 若采用PS加权，应明确说明所选权重类型并提供选用该权重的依据 (7) 若开展基于PS的分层分析，需合理解释所选分层的依据 (8) 在PS匹配队列中进行分析时，需明确说明所采用的是条件效应估计（如使用带匹配对随机效应的混合模型）还是边际效应估计（如使用稳健标准误的回归模型），并注意其解释差异 (9) 应同时开展基于原始未匹配样本、经变量调整的多变量回归分析 (10) 建议使用不同方法验证结果稳健性，包括使用不同PS方法、采用定量工具（如E值）评估未测量混杂对研究结论可能产生的影响等
结果部分	(1) 详细报告应用PS方法调整前后各队列的样本量 (2) 详细报告原始队列及经PS方法调整后，组间协变量的平衡情况 (3) 呈现PS及权重（若使用）的分布特征 (4) 详细报告基于原始数据以及所采用PS方法的效应估计值，并附精确度（如95%置信区间） (5) 详细报告敏感性分析结果
讨论部分	(1) 在对结果进行因果关联解读时，应采取审慎态度，并系统说明分析过程中可能存在的局限，例如匹配后依然存在的协变量残余不平衡、样本量损失对统计效能的影响等 (2) 需明确指出，PS分析不能替代能控制未测量混杂的随机对照试验

量，基于学科知识与因果假设选择协变量构建PS模型，并详细说明模型的构建细节；其次，需报告检验PS关键假设的方法；随后，需明确阐述所采用PS方法的核心参数设置及其选用依据；最后，需报告因果效应估计策略，并说明为验证结果稳健性所采用的敏感性分析策略。

在结果部分，应系统展示应用PS方法调整前后的样本量、协变量平衡情况、PS/权重的分布，并呈现基于原始数据与PS方法的效应估计值以及敏感性分析结果。

最后，在解释研究结果时应保持审慎态度，说明分析中可能存在的局限，不宜过度推论其因果含义。

3 PS的局限与最新进展

尽管PS方法能够通过匹配、分层、加权或回归等方式改善暴露组与对照组之间已测量协变量的平衡，从而降低混杂偏倚，但其结果解释仍需保持审慎。首先，PS方法主要用于平衡已测量协变量，对于未测量或测量不充分的混杂因素仍难以有效控制，因此即使平衡性诊断结果良好，也不能认为残余混杂已被完全消除^[21]。其次，PS分析的效果在较大程度上依赖于协变量选择、变

量处理方式及模型设定，若关键混杂因素遗漏或模型构建不当，仍可能影响效应估计的准确性；此外，在真实世界研究中，治疗可能随时间推移而发生变化，因此PS模型的适用性并非固定不变，在时间跨度较长或数据持续更新的研究中，仍需结合临床背景对模型进行适当调整^[21]。再次，PS方法的有效应用依赖于暴露组与对照组之间具有足够的PS分布重叠，若组间重叠不足，则匹配或加权后所得结果的稳定性和可解释性均会受到影响，尤其在加权分析中，PS接近0或1时还可能产生极端权重，降低估计精度^[53]。最后，PS方法主要用于估计总体平均处理效应，对于不同人群中治疗异质性的检验与估计相对不足；若直接基于整体样本构建PS模型并开展亚组分析，可能无法保证各亚组内协变量达到良好平衡，从而影响异质性效应估计的准确性^[54]。

近年来，PS的方法学不断发展，在传统逻辑回归估计PS的基础上，逐步形成了以高维PS、双重稳健估计和机器学习PS为代表的扩展方法。高维PS主要面向行政数据库和电子病历等高维真实世界数据，通过半自动化算法从大量候选变量中识别并优先纳入可能影响暴露分配和结局风险的变量，从而进一步提高对混杂因素的控制能

力^[55]。双重稳健估计通过同时构建 PS 模型与结局回归模型实现,其核心优势在于:只要任一模型设定正确,即可获得效应量的一致估计,因此在观察性研究复杂资料分析中表现出较高的统计稳健性^[56]。机器学习 PS 则借助广义提升模型 (generalized boosted regression models, GBM)、分类与回归树 (classification and regression tree, CART)、随机森林及神经网络等方法,对非线性关系和高阶交互作用进行更灵活建模,特别适用于高维、复杂数据结构下的 PS 估计^[57]。

4 结语

本文结合具体案例,对指南第 2 版中关于 PS 的内容进行深入解读,旨在为临床与流行病学研究者提供规范的 PS 应用与报告实务参考。在使用 PS 时,研究者应依托清晰的因果推断框架,审慎选择协变量,合理构建模型,并对关键假设和结果的稳健性进行系统评估。

利益冲突声明: 作者声明本研究不存在任何经济或非经济利益冲突。

参考文献

- 1 曾繁典,郑荣远,詹思延,等主编. 药物流行病学,第 2 版[M]. 北京:中国医药科技出版社,2016: 1-650.
- 2 颜济南,吴昉效,聂晓璐,等. 《中国药物流行病学研究方法学指南(第 2 版)》的制订/修订过程[J]. 药物流行病学杂志,2025,34(2): 121-35. [Yan JN, Wu YX, Nie XL, et al. Revision process of the *Guide on Methodological Standards in Pharmacoepidemiology in China (2nd edition)*[J]. Chinese Journal of Pharmacoepidemiology, 2025, 34(2): 121-135.] DOI: 10.12173/j.issn.1005-0698.202502028.
- 3 中国药学会. 中国药学会关于发布《药物流行病学研究方法学指南(第 2 版)》团体标准的公告[EB/OL]. (2025-11-03) [2026-03-16]. <https://www.cpa.org.cn/index.php?do=info&cid=76925>.
- 4 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects[J]. Biometrika, 1983, 70(1): 41-55. DOI: 10.1093/biomet/70.1.41.
- 5 Chen JW, Maldonado DR, Kowalski BL, et al. Best practice guidelines for propensity score methods in medical research: consideration on theory, implementation, and reporting. A review[J]. Arthroscopy, 2022, 38(2): 632-642. DOI: 10.1016/j.arthro.2021.06.037.
- 6 Jackson JW, Schmid I, Stuart EA. Propensity scores in pharmacoepidemiology: beyond the horizon[J]. Curr Epidemiol Rep, 2017, 4(4): 271-280. DOI: 10.1007/s40471-017-0131-y.
- 7 Saarela O, Stephens DA, Moodie EE. The role of exchangeability in causal inference[J]. Stat Sci, 2023, 38(3): 369-385. DOI: 10.1214/22-STS879.
- 8 Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models[J]. Am J Epidemiol, 2008, 168(6): 656-664. DOI: 10.1093/aje/kwn164.
- 9 Tennant PWG, Murray EJ, Arnold KF, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations[J]. Int J Epidemiol, 2021, 50(2): 620-632. DOI: 10.1093/ije/dyaa213.
- 10 韦柏安,陈耿杭,刘少南,等. 有向无环图构建与应用步骤解读[J]. 中国循证医学杂志,2025,25(11): 1350-1357. [Wei BA, Chen GH, Liu SN, et al. Interpretation of the steps in the construction and application of directed acyclic graphs[J]. Chinese Journal of Evidence-Based Medicine, 2025, 25(11): 1350-1357.] DOI: 10.7507/1672-2531.202506078.
- 11 Ding P, VanderWeele TJ, Robins JM. Instrumental variables as bias amplifiers with general outcome and confounding[J]. Biometrika, 2017, 104(2): 291-302. DOI: 10.1093/biomet/asx009.
- 12 Tönnies T, Kahl S, Kuss O. Collider bias in observational studies[J]. Dtsch Arztebl Int, 2022, 119(7): 107-122. DOI: 10.3238/arztebl.m2022.0076.
- 13 van Zwieten A, Tennant PWG, Kelly-Irving M, et al. Avoiding overadjustment bias in social epidemiology through appropriate covariate selection: a primer[J]. J Clin Epidemiol, 2022, 149: 127-136. DOI: 10.1016/j.jclinepi.2022.05.021.
- 14 VanderWeele T. Principles of confounder selection[J]. Eur J Epidemiol, 2019, 34(3): 211-219. DOI: 10.1007/s10654-019-00494-6.
- 15 Hoffman KL, Schenck EJ, Satlin MJ, et al. Comparison of a target trial emulation framework vs cox regression to estimate the association of corticosteroids with COVID-19 mortality[J]. JAMA Netw Open, 2022, 5(10): e2234425. DOI: 10.1001/jamanetworkopen.2022.34425.
- 16 Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and Meta-classifiers as alternatives to Logistic regression[J]. J Clin Epidemiol, 2010, 63(8): 826-833. DOI: 10.1016/j.jclinepi.2009.11.020.
- 17 Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology[J]. Epidemiology, 2000, 11(5): 550-560. DOI: 10.1097/00001648-200009000-00011.
- 18 Austin P. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples[J]. Stat Med, 2009, 28(25): 3083-3107. DOI: 10.1002/sim.3697.
- 19 Stuart EA, Lee BK, Leacy FP. Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research[J]. J Clin Epidemiol, 2013, 66(8 Suppl): S84-S90.e1. DOI: 10.1016/j.jclinepi.2013.01.013.
- 20 Lee J, Little TD. A practical guide to propensity score analysis for applied clinical research[J]. Behav Res Ther, 2017, 98: 76-90.

- DOI: [10.1016/j.brat.2017.01.005](https://doi.org/10.1016/j.brat.2017.01.005).
- 21 Loke YK, Mattishent K. Propensity score methods in real-world epidemiology: a practical guide for first-time users[J]. *Diabetes Obes Metab*, 2020, 22 Suppl 3: 13–20. DOI: [10.1111/dom.13926](https://doi.org/10.1111/dom.13926).
 - 22 Imbens GW. Nonparametric estimation of average treatment effects under exogeneity: a review[J]. *Rev Econ Stat*, 2004, 86(1): 4–29. DOI: [10.3386/t0294](https://doi.org/10.3386/t0294).
 - 23 Matsouka RA, Zhou Y. Causal inference in the absence of positivity: the role of overlap weights[J]. *Biom J*, 2024, 66(4): e2300156. DOI: [10.1002/bimj.202300156](https://doi.org/10.1002/bimj.202300156).
 - 24 Greifer N, Stuart EA. Choosing the estimand when matching or weighting in observational studies[EB/OL]. (2021–06–19) [2026–02–28]. <https://arxiv.org/abs/2106.10577v1>.
 - 25 Stuart EA. Matching methods for causal inference: a review and a look forward[J]. *Stat Sci*, 2010, 25(1): 1–21. DOI: [10.1214/09-sts313](https://doi.org/10.1214/09-sts313).
 - 26 Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score[J]. *Am Stat*, 1985, 39(1): 33–38. <https://doi.org/10.2307/2683903>.
 - 27 Rosenbaum PR. Optimal matching for observational studies[J]. *J Am Stat Assoc*, 1989, 84(408): 1024–1032. <https://doi.org/10.2307/2290079>.
 - 28 Austin PC. A comparison of 12 algorithms for matching on the propensity score[J]. *Stat Med*, 2014, 33(6): 1057–1069. DOI: [10.1002/sim.6004](https://doi.org/10.1002/sim.6004).
 - 29 Stuart EA, Green KM. Using full matching to estimate causal effects in nonexperimental studies: examining the relationship between adolescent marijuana use and adult outcomes[J]. *Dev Psychol*, 2008, 44(2): 395–406. DOI: [10.1037/0012-1649.44.2.395](https://doi.org/10.1037/0012-1649.44.2.395).
 - 30 Austin PC, Stuart EA. The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes[J]. *Stat Methods Med Res*, 2017, 26(4): 1654–1670. DOI: [10.1177/0962280215584401](https://doi.org/10.1177/0962280215584401).
 - 31 Glimm E, Yau L. Exact matching as an alternative to propensity score matching[J/OL]. *Stat Biopharm Res*, 2025: 1–11. [2026–02–28]. DOI: [10.48550/arXiv.2503.02850](https://doi.org/10.48550/arXiv.2503.02850).
 - 32 King G, Nielsen R. Why propensity scores should not be used for matching[J]. *Polit Anal*, 2019, 27(4): 435–454. DOI: [10.1017/pan.2019.11](https://doi.org/10.1017/pan.2019.11).
 - 33 Ramsey DJ, Makwana B, Dani SS, et al. GLP-1 receptor agonists and sight-threatening ophthalmic complications in patients with type 2 diabetes[J]. *JAMA Netw Open*, 2025, 8(8): e2526321. DOI: [10.1001/jamanetworkopen.2025.26321](https://doi.org/10.1001/jamanetworkopen.2025.26321).
 - 34 Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies[J]. *Multivariate Behav Res*, 2011, 46(3): 399–424. DOI: [10.1080/00273171.2011.568786](https://doi.org/10.1080/00273171.2011.568786).
 - 35 Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies[J]. *Biometrics*, 1968, 24(2): 295–313. DOI: [10.2307/2528036](https://doi.org/10.2307/2528036).
 - 36 Desai RJ, Rothman KJ, Bateman BT, et al. A propensity-score-based fine stratification approach for confounding adjustment when exposure is infrequent[J]. *Epidemiology*, 2017, 28(2): 249–257. DOI: [10.1097/ede.0000000000000595](https://doi.org/10.1097/ede.0000000000000595).
 - 37 Elze MC, Gregson J, Baber U, et al. Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies[J]. *J Am Coll Cardiol*, 2017, 69(3): 345–357. DOI: [10.1016/j.jacc.2016.10.060](https://doi.org/10.1016/j.jacc.2016.10.060).
 - 38 Schafer JL, Kang J. Average causal effects from nonrandomized studies: a practical guide and simulated example[J]. *Psychol Methods*, 2008, 13(4): 279–313. DOI: [10.1037/a0014268](https://doi.org/10.1037/a0014268).
 - 39 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: 15657. DOI: [10.1136/bmj.l5657](https://doi.org/10.1136/bmj.l5657).
 - 40 Kurz CF, Krzywinski M, Altman N. Propensity score weighting[J]. *Nat Methods*, 2025, 22(4): 638–640. DOI: [10.1038/s41592-025-02629-y](https://doi.org/10.1038/s41592-025-02629-y).
 - 41 Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies[J]. *Stat Med*, 2015, 34(28): 3661–3679. DOI: [10.1002/sim.6607](https://doi.org/10.1002/sim.6607).
 - 42 Brookhart MA, Wyss R, Layton JB, et al. Propensity score methods for confounding control in nonexperimental research[J]. *Circ Cardiovasc Qual Outcomes*, 2013, 6(5): 604–611. DOI: [10.1161/circoutcomes.113.000359](https://doi.org/10.1161/circoutcomes.113.000359).
 - 43 Yoshida K, Hernández-Díaz S, Solomon DH, et al. Matching weights to simultaneously compare three treatment groups: comparison to three-way matching[J]. *Epidemiology*, 2017, 28(3): 387–395. DOI: [10.1097/ede.0000000000000627](https://doi.org/10.1097/ede.0000000000000627).
 - 44 Li F, Thomas LE, Li F. Addressing extreme propensity scores via the overlap weights[J]. *Am J Epidemiol*, 2019, 188(1): 250–257. DOI: [10.1093/aje/kwy201](https://doi.org/10.1093/aje/kwy201).
 - 45 Li F. Propensity score weighting for causal inference with multi-valued treatments[EB/OL]. (2018–08–16) [2026–02–28]. <https://arxiv.org/abs/1808.05339>.
 - 46 Xu KY, Hartz SM, Borodovsky JT, et al. Association between benzodiazepine use with or without opioid use and all-cause mortality in the United States, 1999–2015[J]. *JAMA Netw Open*, 2020, 3(12): e2028557. DOI: [10.1001/jamanetworkopen.2020.28557](https://doi.org/10.1001/jamanetworkopen.2020.28557).
 - 47 Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study[J]. *Stat Med*, 2004, 23(19): 2937–2960. DOI: [10.1002/sim.1903](https://doi.org/10.1002/sim.1903).
 - 48 Staffa SJ, Zurawski D. Five steps to successfully implement and evaluate propensity score matching in clinical research studies[J]. *Anesth Analg*, 2018, 127(4): 1066–1073. DOI: [10.1213/ane.0000000000002787](https://doi.org/10.1213/ane.0000000000002787).
 - 49 Nguyen TL, Collins GS, Spence J, et al. Double-adjustment in

- propensity score matching analysis: choosing a threshold for considering residual imbalance[J]. *BMC Med Res Methodol*, 2017, 17(1): 78. DOI: [10.1186/s12874-017-0338-0](https://doi.org/10.1186/s12874-017-0338-0).
- 50 VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value[J]. *Ann Intern Med*, 2017, 167(4): 268–274. DOI: [10.7326/m16-2607](https://doi.org/10.7326/m16-2607).
- 51 Li L, Shen C, Wu AC, et al. Propensity score-based sensitivity analysis method for uncontrolled confounding[J]. *Am J Epidemiol*, 2011, 174(3): 345–353. DOI: [10.1093/aje/kwr096](https://doi.org/10.1093/aje/kwr096).
- 52 Andrew BY, Alan Brookhart M, Pearce R, et al. Propensity score methods in observational research: brief review and guide for authors[J]. *Br J Anaesth*, 2023, 131(5): 805–809. DOI: [10.1016/j.bja.2023.06.054](https://doi.org/10.1016/j.bja.2023.06.054).
- 53 Zhou Y, Matsouaka RA, Thomas L. Propensity score weighting under limited overlap and model misspecification[J]. *Stat Methods Med Res*, 2020, 29(12): 3721–3756. DOI: [10.1177/0962280220940334](https://doi.org/10.1177/0962280220940334).
- 54 Chatelet F, Verillaud B, Chevret S. How to perform prespecified subgroup analyses when using propensity score methods in the case of imbalanced subgroups[J]. *BMC Med Res Methodol*, 2023, 23(1): 255. DOI: [10.1186/s12874-023-02071-8](https://doi.org/10.1186/s12874-023-02071-8).
- 55 Karim ME. High-dimensional propensity score and its machine learning extensions in residual confounding control[J]. *Am Stat*, 2025, 79(1): 72–90. DOI: [10.1080/00031305.2024.2368794](https://doi.org/10.1080/00031305.2024.2368794).
- 56 Funk MJ, Westreich D, Wiesen C, et al. Doubly robust estimation of causal effects[J]. *Am J Epidemiol*, 2011, 173(7): 761–767. DOI: [10.1093/aje/kwq439](https://doi.org/10.1093/aje/kwq439).
- 57 Leite W, Zhang H, Collier Z, et al. Machine learning for propensity score estimation: a systematic review and reporting guidelines[J/OL]. *Psychol Methods*, 2025. [2026-02-28]. DOI: [10.1037/met0000789](https://doi.org/10.1037/met0000789).

收稿日期: 2026 年 03 月 09 日 修回日期: 2026 年 03 月 20 日
 本文编辑: 洗静怡 杨 燕