

· 综述 ·

观察性研究中的因果图和混杂因素的识别和选择准则



刘慧珍¹, 周小芹¹, 王 婷¹, 康德英^{1, 2}

1. 四川大学华西医院临床研究管理部临床方案设计与统计办公室 (成都 610041)

2. 四川大学华西医院临床流行病学与循证医学研究中心 (成都 610041)

【摘要】 在观察性研究中, 准确可靠地估计暴露对结局的因果效应是其核心目标, 而确定并充分矫正混杂因素是实现这一目标的关键前提和核心挑战。混杂因素控制不当——无论是遗漏关键混杂 (导致残余混杂), 还是过度调整无关变量 (引入碰撞偏倚等)——都会直接扭曲效应估计结果, 进而可能产生误导性的科学结论和临床决策。因此, 开发并应用系统、透明且可重复的混杂因素识别与选择策略, 对提升观察性研究因果推断的效度与可靠性具有根本性意义。本文通过系统综述有向无环图 (DAGs) 这一强大的可视化因果建模工具, 并深入评析 3 种基于 DAGs 的主流混杂因素选择准则 (Pre-exposure 准则、Common cause 准则、Modified disjunctive cause 准则), 为研究者提供一个结构化、理论依据充分的混杂因素识别与选择框架, 以期显著优化观察性研究中因果效应的估计过程。

【关键词】 混杂因素; 混杂; 有向无环图; Pre-exposure 准则; Common cause 准则; Modified disjunctive cause 准则

【中图分类号】 R 181.3+5

【文献标识码】 A

Causal diagrams and criteria for identifying and selecting confounders in observational studies

LIU Huizhen¹, ZHOU Xiaoqin¹, WANG Ting¹, KANG Deying^{1,2}

1. Center of Biostatistics, Design, Measurement and Evaluation (CBDME), Department of Clinical Research Management, West China Hospital, Sichuan University, Chengdu 610041, China

2. Department of Evidence-Based Medicine and Clinical Epidemiology, West China Hospital, Sichuan University, Chengdu 610041, China

Corresponding author: KANG Deying, Email: deyingkang12@163.com

【Abstract】 In observational research, a primary objective is to accurately and reliably assess the causal impact of exposure on outcomes. Identifying and properly adjusting for confounding factors is a key prerequisite and central challenge to achieving this goal. Ineffective management of confounders, whether by neglecting significant ones, (leading to residual confounding), or by over-adjusting for irrelevant factors, (introducing collider bias), can distort effect estimates and lead to erroneous scientific conclusions and clinical decisions. Therefore, it is essential to develop and implement systematic, transparent, and reproducible methods for identifying and selecting confounding factors to enhance the validity and reliability of causal

DOI: 10.12173/j.issn.1005-0698.202503160

基金项目: 四川大学华西医院学科卓越发展 1·3·5 工程项目 (ZYGD23002)

通信作者: 康德英, 教授, 硕士研究生导师, Email: deyingkang12@163.com

inferences in observational studies. This paper provides a systematic review of directed acyclic graphs (DAGs), a robust visual causal modeling tool, and offers a detailed examination of three prominent criteria for selecting confounding factors based on DAGs: the Pre-exposure criterion, the Common cause criterion, and the Modified disjunctive cause criterion. The aim is to equip researchers with a structured and theoretically grounded framework for identifying and selecting confounding factors, thereby improving the process of estimating causal effects in observational studies.

【Keywords】 Confounders; Confounding; Directed acyclic graphs; Pre-exposure criterion; Common cause criterion; Modified disjunctive cause criterion

开展随机对照试验存在伦理约束、研究成本（时间、经济等）较高、受试者纳排标准严苛，受试者招募困难等问题^[1-3]。伴随着医学大数据的普及，观察性研究在医学研究中发挥着重要作用^[4]。在观察性研究中探究因果关系时，若暴露组与对照组间的结局差异源于非暴露因素时，即发生混杂^[5]。若在研究设计或分析阶段未控制混杂因素，会扭曲暴露与结局间的真实关联，进而误导临床决策^[6-7]。混杂因素指同时与暴露和结局均有关系的因素^[5, 8-9]。在观察性研究中，研究者通常在设计阶段通过限制、分层，或在分析阶段通过回归模型、匹配等方法对其进行控制^[10]。在观察性研究中，选择正确的混杂因素对因果推断至关重要。若遗漏关键混杂因素会导致研究结果存在残余混杂；若控制无关变量则可能引发“过度调整”，造成标准误膨胀、估计偏倚（如碰撞偏倚），或两者兼具^[8]。当前，潜在混杂因素的选择多依赖专业知识，相关决策过程常缺乏透明度。

因果推断理论的发展为混杂因素识别提供了新的方法论支持，而有向无环图（directed acyclic graphs, DAGs）作为可视化因果结构的工具，已成为当前研究的热点方向^[11-13]。其通过节点变量间的定向边连接，直观呈现暴露因素、结局变量及潜在混杂因素间的拓扑关系，能辅助研究者基于先验知识构建因果模型，进而提升混杂因素识别的系统性和透明度。然而，一项针对 234 篇使用 DAGs 的文章的回顾性分析^[14]发现，DAGs 的使用与报告仍存较大差异。在混杂因素选择准则方面，近年来涌现出多种创新性方法^[11]。Pre-exposure 准则通过时间顺序确保混杂因素在暴露前确定，提高了因果推断的时间合理性；Common cause 准则聚焦识别暴露与结局的共同原因，避免了无关变量的控制；Disjunctive cause 准则及其

改进版本 Modified disjunctive cause 准则则通过辨别暴露或结局的原因，进一步优化了混杂因素的选择效率。这些准则为研究者提供了更客观的选择依据。

然而，当前相关研究仍存在一些亟待解决的问题：① DAGs 的构建高度依赖先验知识，不同研究者可能得出不同的因果结构；② 现有混杂因素选择准则各有侧重，尚缺乏统一的适用标准。本文通过系统梳理 DAGs 在因果路径可视化及辅助混杂识别中的应用原理，同时全面整合并比较当前主流混杂因素选择准则，提出更具操作性的混杂因素识别框架；此外，本文还对不同准则的适用场景进行比较分析，为研究者提供方法学选择的决策依据，帮助研究者解决面对多种准则时的选择困惑，提升混杂因素选择过程的透明度、规范性与可重复性；同时有助于加强研究者对混杂因素选择逻辑的理解，辅助其更好地报告与解释流行病学研究结果，进而显著提高观察性研究中因果推断的透明度和效度。

1 DAGs简介

DAGs 是一种可视化变量间因果关系的图形工具，能帮助研究者识别研究中的关键混杂因素，进而指导变量选择^[11]。在 DAGs 中，节点表示变量，连接各节点的有向边（即箭头，箭头方向代表时序为由因到果）代表变量间的因果关系。通过箭头连接各节点，DAGs 可形成定向路径，但不会产生闭合回路^[11]。一个完整的因果 DAGs 需包含从暴露到结局路径上的所有可能变量，以及可能对这些变量产生因果影响的其他变量，且所有变量需通过一系列同向单向箭头连接而成。DAGs 的构建以研究者对系统内变量间因果关系的理论假设为基础。这些假设可来源于先验知识、文献综述、探索性分析结果等。

DAGs 包含 3 种基本路径^[12, 15], 分别为链状结构、叉状结构和碰撞结构, 见图 1。在链状结构中, 3 个节点间的 2 条有向边, 1 条由暴露因素指向中间变量, 另 1 条由中间变量指向结局变量, 此时的中间变量为中介变量。在叉状结构中, 3 个节点间的 2 条有向边均由中间变量发出, 分别指向暴露变量与结局变量, 此时的中间变量为混杂变量。在碰撞结构中, 3 个节点间的 2 条有向边分别由暴露变量、结局变量指向中间变量, 此时的中间变量为碰撞变量。若 2 个节点间存在链状结构或叉状结构时, 该路径即为开放路径。根据有向联接规则, 开放路径仅包括链状结构时, 则该路径为病因路径; 若 2 个节点间包括叉状结构时, 该路径为混杂路径, 由该路径产生的统计学关联为虚假关联。开放路径表示 2 个变量可能存在非独立性, 即相关性。若 2 个节点间存在碰撞结构时, 该路径为阻断路径。根据有向分离规则, 阻断路径上碰撞变量两侧的变量相互独立, 不存在因果关联。矫正阻断路径中的非碰撞节点会阻断该路径, 而矫正阻断路径中的碰撞节点则会开放此路径。

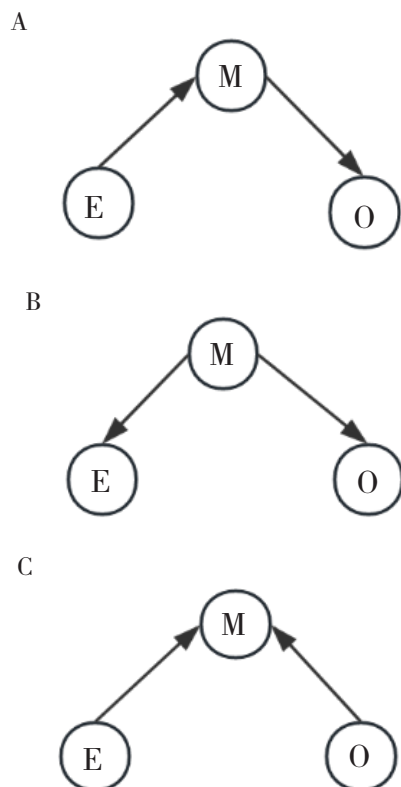


图1 DAGs中3种基本路径图

Figure 1. Three fundamental path diagrams in a DAGs

注: A. 链状结构; B. 叉状结构; C. 碰撞结构; E. 暴露因素; O. 结局变量; M. 中间变量。

根据 DAGs 的逻辑, 观察性研究中矫正混杂因素也就是阻断所有开放的后门路径。阻断后门路径主要方式有 2 种方式^[16]: ①采用限制、匹配或分层的方式控制需矫正的混杂因素, 见图 2-A。该方法是将研究人群按照混杂因素进行分层 (设计阶段可采用限制或匹配, 统计分析阶段按照分层的方式), 使每层内暴露组与非暴露组内的混杂因素分布均衡。随后, 分别估计每层内暴露因素与研究结局之间的关联程度 [如比值比 (odds Rratio, OR) 或相对危险度 (relative risk, RR)], 再通过 Mantel-Haenszel 等分层分析方法, 综合各层的 OR 或 RR, 最终得到矫正混杂因素后的调整 OR 或 RR。②删除混杂因素到暴露因素的路径。采用标准化、逆概率加权等方式删除混杂因素与暴露因素之间的路径, 使暴露因素与混杂因素不再相关, 即暴露因素对结局变量的影响不受混杂因素干扰, 见图 2-B。当暴露因素不是时依变量时, 以上 2 种阻断方式均可矫正研究中的混杂因素, 但若暴露因素随时间变化时, 则使用如 G-estimation、边际结构模型等方法矫正研究中的混杂因素。

DAGs 能将研究中的各种复杂关系可视化, 帮助研究者直观识别混杂因素, 但在实际应用中仍存在一定限制^[16-17]。首先, DAGs 的绘制高度依赖研究者专业领域的知识储备。面对复杂的因果关系网络时, 常可能生成多个候选 DAGs, 此时难以直接判断最优选。因此需穷尽所有潜在 DAGs, 通过确定每个 DAGs 中的最小矫正子集,

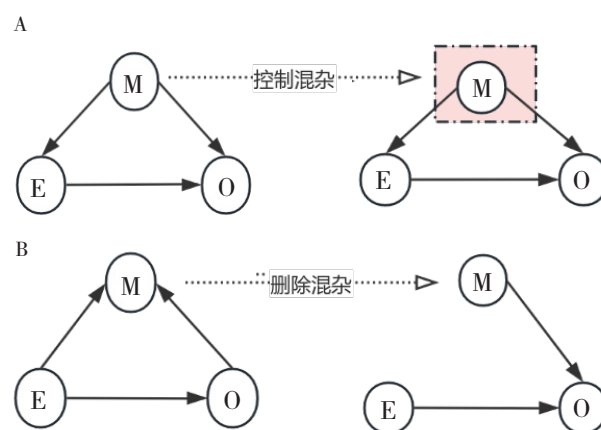


图2 基于DAGs进行混杂因素识别和控制

Figure 2. Recognizing and managing confounding variables using DAGs

注: A. 控制混杂; B. 删除混杂; E. 暴露因素; O. 结局变量; M. 混杂变量。

才能确保结论的可靠性。第 2 个限制源于不可观测的混杂变量，即便经过最优调整，研究中仍可能存在残留混杂，这导致基于 DAGs 的因果效应估计未必完全无偏，但能显著提升估计值的近似精度。为充分发挥该工具的优势，研究者可采取双重优化策略：一方面深化领域认知，构建更贴近真实因果关系的 DAGs；另一方面在研究设计阶段，对已识别的混杂因子进行系统测量，从而有效控制残余混杂。通过这种双重优化策略，可最大限度确保因果效应估计值逼近真实参数。

2 混杂因素选择准则

2.1 Pre-exposure 准则

Pre-exposure 准则最早由 Rubin^[18] 提出，部分文献也将其称为 Pre-treatment 准则^[19]。该准则的理论依据：所有同时影响暴露 / 治疗因素与结局的混杂变量，本质上都是暴露 / 治疗前已存在的干扰因素，因此强调仅需对所有可获得的暴露 / 治疗前变量进行调整^[20]。由于部分变量对暴露 / 治疗因素与结局的混杂作用机制尚未明确，基于混杂因素时序性调整的原则，矫正全部暴露 / 治疗前变量，不失为一种稳健策略。这种方法选择源于对因果推断中混杂结构不确定性的审慎考量——通过优先控制暴露 / 治疗前所有潜在混杂源，可系统性降低因遗漏混杂或误判混杂方向导致的因果效应估计偏倚。然而，若被调整的变量属于碰撞变量 (collider)，可能导致对撞 / 碰撞偏倚 (collider bias 或 M-bias)^[21]。如在子代肥胖与子代糖尿病关系研究中 (图 3)，若矫正了碰撞变量 U (母亲患糖尿病)，即在 U 的某一取值水平上估计暴露 E (子代肥胖) 与结局 O (子代

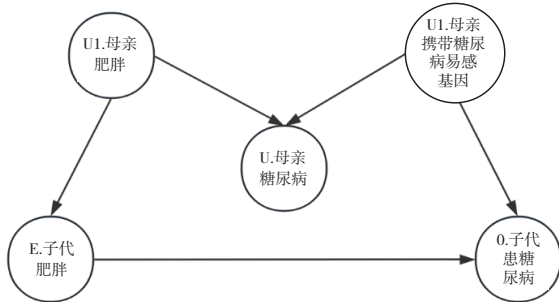


图3 研究调整因素为碰撞变量举例

Figure 3. An example of adjustment factor as a collision variable

注：E. 暴露变量；O. 结局变量；U. 碰撞变量。

糖尿病) 的关联，会开启暴露因素到研究结局间的后门路径，进而扭曲两者的真实关系。

2.2 Common cause 准则

Common cause 准则可绕开对撞 / 碰撞偏倚 (collider bias 或 M-bias)，其核心内容为在研究某一暴露因素对健康结局的影响时，必须找出并控制所有研究开始前已存在且同时影响暴露变量与结局变量的因素^[22]。该准则基于因果图的先验知识，需判明各变量是否同时影响暴露变量和结局变量 (相较于基于 DAGs 所需的后门准则，该准则所需的因果链条知识大幅简化)。根据 Common cause 准则，若对所有暴露与结局的共同诱因进行控制，无论其因果结构如何，均可有效矫正暴露对结局的混杂效应。但当某些共同诱因不可获取或数据缺失时，按照该准则则不能调整，此时研究中可能会存在未调整的混杂偏倚。在这种情况下，可通过 DAGs 识别一个不是共同原因的变量，对其进行调整，仍可消除暴露和结局的共同原因的混杂效应。如图 4 所示，未测量的混杂 U (基线免疫状态) 为暴露因素 E (抗菌药物使用) 和结局变量 O (术后感染风险) 的共同原因，但该变量未被观察，变量 C (白细胞计数) 仅为暴露因素 (抗菌药物使用) 的原因，按照 Common cause 准则需要矫正 U，而不是 C。然而 U 不可测或数据缺失，此时可通过调整变量 C 来消除混杂因素的干扰。

2.3 Modified disjunctive cause 准则

Disjunctive cause 准则要求调整暴露前的变量，这些变量可能是暴露的原因、结局的原因，也可能是暴露和结局共同的原因^[23]。Disjunctive cause 准则避免选择碰撞变量 (如图 3 中的 U)，同时可控制影响暴露的变量 (若暴露变量不可测时，可通过调整影响暴露的变量来控制混杂，如图 4 中的 C)，因此 Disjunctive cause 准则克服了 Pre-exposure 准则 (可能调整非必要的变量，如碰撞变量) 和 Common cause 准则 (可能未调整本可以消除偏差的混杂因素) 的局限性。与 Common cause 准则类似，使用 Disjunctive cause 准则需要明确每个变量是否影响暴露或结局。VanderWeele 等^[23] 的研究表明，基于 Disjunctive cause 准则选择协变量可有效调整混杂。

在流行病学研究的混杂控制中，Disjunctive cause 准则为协变量筛选提供了系统化解方案

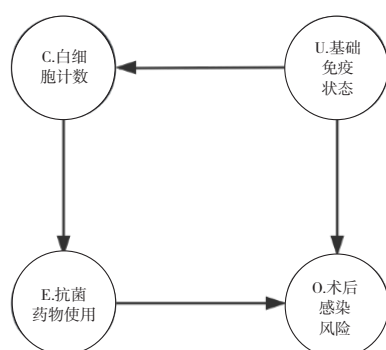


图4 研究混杂因素为不可测量因素举例

Figure 4. An example of an unmeasurable factor as a confounding variable in research

注：E. 暴露变量；O. 结局变量；U. 不可测量变量。

案，尤其适用于混杂因素与暴露、结局间因果关系明确的场景。然而，该准则在应用时也存在一定的局限性。具体而言，若控制仅影响暴露的工具变量可能因存在未测量混杂因素而意外放大偏倚，此现象称为 Z 偏倚 (Z-bias) [24]。工具变量是暴露因素的原因，但对结局没有直接的因果关系，也不能通过暴露以外的其他变量间接影响结局事件 [25]。工具变量可用于估计因果效应，但在回归模型或倾向性评分模型中控制工具变量可能会产生额外的偏差。如探讨身体活动 (暴露) 与 2 型糖尿病发病率 (结局事件) 之间的因果关系研究中，若经研究证明某基因位点仅通过增加体力活动水平影响糖尿病发病风险，此时矫正该基因位点会间接放大非测量的混杂路径 (如社会经济地位较低者可能会同时携带该特定基因型且体力活动不足)，从而错误估计体力活动与糖尿病发病风险之间的因果效应 (图 5)。因此，应用 Disjunctive cause 准则时需额外排除工具变量。

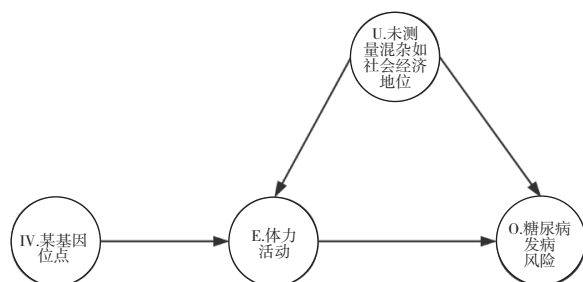


图5 研究控制工具变量的举例

Figure 5. An example of controlling for an instrumental variable in research

注：E. 暴露变量；O. 结局变量；IV. 工具变量。

此外，在观察性研究中，控制变量 U (尤其当变量 U 为不可测量变量) 的代理变量 V，有助于阻断后门路径，从而减少研究中的混杂偏倚。该变量应是暴露或结局变量的代理变量，而不是暴露变量和结局变量的共同原因。以长期使用非甾体抗炎药 (nonsteroidal antiinflammatory drugs, NSAIDs, 即暴露因素) 与老年人认知功能下降的关联性研究为例，慢性疼痛严重程度为难以精确量化的因素 (不可测量因素 U)，其可能同时影响 NSAIDs 的使用 (疼痛严重者更可能长期用药) 与认知功能 (慢性疼痛可通过炎症通路损害认知)。在临床研究中，可使用门诊疼痛类药物处方频率作为 U 的代理变量 V，矫正 V 后，能阻断 “U → NSAIDs 的使用 → 老年人认知功能下降” 这一后门路径，进而部分控制未测量的疼痛严重程度对研究结果的干扰 (图 6)。若代理变量为暴露和结局的共同原因，则会形成类似图 3 的结构，可能导致研究中引入 M-bias，因此在研究设计中，应尽量避免纳入这类作为暴露和结局事件共同原因的代理变量。

随后，VanderWeele 提出了 Modified disjunctive cause 准则 [26]，该准则将协变量选择标准归纳以下 3 个方面的内容：① 纳入所有暴露因素、结局变量或两者的原因变量；② 排除暴露与结局的工具变量；③ 当暴露与结局存在未测量的共同混杂因素时，纳入其代理变量。Modified disjunctive cause 准则可系统识别充分混杂因素集，被学界视为部分指定因果 DAGs 中最全面的混杂控制方法之一 [25]。该准则的主要优势在于其大幅简化了协变量的选择过程。相较于 Common cause 准则，该准则具有更强的可操作性。具体而言，判断某变量是否为暴露或结局的单向原因变量，远比验证其是否同时作用于暴露与结局更为简便。这一特性使该准则尤其适用于复杂流行病学研究中的混杂控制。然而值得注意的是，Modified disjunctive cause 准则的应用仍可能对研究者带来显著挑战。以工具变量筛选为例，其识别过程可能面临较大困难。因鲜有工具变量具有明确的外源性特征，且多数潜在工具变量常与研究中的其他变量存在复杂交互关系 [11]。在此类情形下，研究者需通过系统性的敏感性分析，对不同协变量调整组合进行验证测试，以评估研究结果的稳健性。

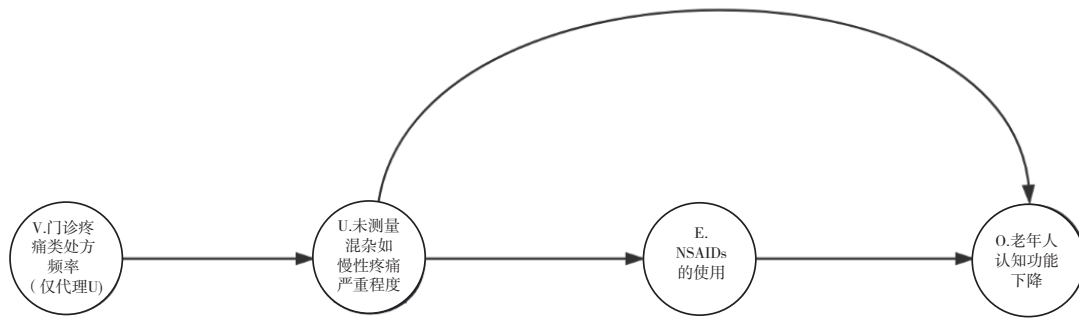


图6 研究控制代理变量的举例

Figure 6. An example of controlling for a proxy variable in research

注：E. 暴露变量；O. 结局变量；V. 代理变量。

3 结语

本文从流行病学和设计角度系统梳理了混杂因素识别体系。基于 DAGs 的因果结构可视化技术，能有效解构暴露 – 结局间的中介路径与混杂网络。Modified disjunctive cause 准则在协变量间因果结构信息有限时仍能实现有效的混杂筛选，其通过整合暴露与结局的潜在共同诱因网络，构建具有操作性的混杂识别框架。随着数据科学的发展，未来可考虑探索如何利用人工智能技术，半自动化地生成或验证因果图，辅助研究者进行混杂因素的识别和选择。

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 Dahabreh IJ, Bibbins-Domingo K. Causal inference about the effects of interventions from observational studies in medical journals[J]. JAMA, 2024, 331(21): 1845–1853. DOI: [10.1001/jama.2024.7741](https://doi.org/10.1001/jama.2024.7741).
- 2 Martin L, Hutchens M, Hawkins C, et al. How much do clinical trials cost?[J]. Nat Rev Drug Discov, 2017, 16(6): 381–382. DOI: [10.1038/nrd.2017.70](https://doi.org/10.1038/nrd.2017.70).
- 3 袁驰, 周祎灵, 曹雨滋, 等. 基于真实世界数据的观察性研究因果推断——目标试验模拟实施要点及案例分析 [J]. 中国胸心血管外科临床杂志, 2024, 31(12): 1743–1752. [Yuan C, Zhou YL, Cao YZ, et al. Causal inference in observational studies based on real-world data: key points and case studies for target trial emulation[J]. Chinese Journal of Clinical Thoracic and Cardiovascular Surgery, 2024, 31(12): 1743–1752.] DOI: [10.7507/1007-4848.202407072](https://doi.org/10.7507/1007-4848.202407072).
- 4 Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials[J]. N Engl J Med, 2000, 342(25): 1878–1886. DOI: [10.1056/NEJM200006223422506](https://doi.org/10.1056/NEJM200006223422506).
- 5 Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research[J]. Epidemiology, 1999, 10(1): 37–48. <https://pubmed.ncbi.nlm.nih.gov/9888278/>.
- 6 Dickerman BA, García-Albéniz X, Logan RW, et al. Avoidable flaws in observational analyses: an application to statins and cancer[J]. Nat Med, 2019, 25(10): 1601–1606. DOI: [10.1038/s41591-019-0597-x](https://doi.org/10.1038/s41591-019-0597-x).
- 7 Hammerton G, Munafò MR. Causal inference with observational data: the need for triangulation of evidence[J]. Psychol Med, 2021, 51(4): 563–578. DOI: [10.1017/S0033291720005127](https://doi.org/10.1017/S0033291720005127).
- 8 Pearl J, Glymour M, Jewell NP. Causal inference in statistics: a primer[M]. New Jersey: John Wiley & Sons, 2016: 1–160.
- 9 VanderWeele TJ, Shpitser I. On the definition of a confounder[J]. Ann Stat, 2013, 41(1): 196–220. DOI: [10.1214/12-aos1058](https://doi.org/10.1214/12-aos1058).
- 10 Inoue K, Sakamaki K, Komukai S, et al. Methodological tutorial series for epidemiological studies: confounder selection and sensitivity analyses to unmeasured confounding from epidemiological and statistical perspectives[J]. J Epidemiol, 2025, 35(1): 3–10. DOI: [10.2188/jea.JE20240082](https://doi.org/10.2188/jea.JE20240082).
- 11 Chan GCK, Sun T, Stjepanović D, et al. Designing observational studies for credible causal inference in addiction research—Directed acyclic graphs, modified disjunctive cause criterion and target trial emulation[J]. Addiction, 2024, 119(6): 1125–1134. DOI: [10.1111/add.16442](https://doi.org/10.1111/add.16442).
- 12 郝允逸, 夏雪, 胥芹, 等. 有向无环图的构建规范: ESC-DAGs 方法介绍及脑血管病研究案例解读 [J]. 中国卒中杂志, 2024, 19(10): 1221–1229. [Hao YY, Xia X, Xu Q, et al. Construction specification of directed acyclic graphs: an introduction to the ESC-DAGs method and case interpretation in cerebrovascular disease research[J]. Chinese Journal of Stroke, 2024, 19(10): 1221–1229.] DOI: [10.3969/j.issn.1673-5765.2024.10.016](https://doi.org/10.3969/j.issn.1673-5765.2024.10.016).
- 13 Feeney T, Hartwig FP, Davies NM. How to use directed acyclic graphs: guide for clinical researchers[J]. BMJ, 2025, 388: e78226. DOI: [10.1136/bmj-2023-078226](https://doi.org/10.1136/bmj-2023-078226).
- 14 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](https://doi.org/10.1093/ije/dyaa213).
- 15 姜弘扬. 面向生物医学数据的因果推断方法研究 [D]. 长春:

- 吉林大学, 2024. DOI: [10.27162/d.cnki.gjlin.2024.007467](https://doi.org/10.27162/d.cnki.gjlin.2024.007467).
- 16 刘慧鑫, 汪海波, 汪宁. 有向无环图在混杂因素识别与控制中的应用及实例分析[J]. 中华流行病学杂志, 2020, 41(4): 585–588. [Liu HX, Wang HB, Wang N. Application of directed acyclic graphs in identifying and controlling confounding bias[J]. Chinese Journal of Epidemiology, 2020, 41(4): 585–588.] DOI: [10.3760/cma.j.cn112338-20190729-00559](https://doi.org/10.3760/cma.j.cn112338-20190729-00559).
- 17 刘子言, 吴小丽, 解美秋, 等. 在因果推断中应用有向无环图识别和控制选择偏倚[J]. 中华疾病控制杂志, 2019, 23(3): 351–355. [Liu ZY, Wu XY, Xie MQ, et al. Application of directed acyclic graphs in identification and control of selection bias in causal inference[J]. Chinese Journal of Disease Control & Prevention, 2019, 23(3): 351–355.] DOI: [10.16462/j.cnki.zhjbkz.2019.03.022](https://doi.org/10.16462/j.cnki.zhjbkz.2019.03.022).
- 18 Rubin DB. Should observational studies be designed to allow lack of balance in covariate distributions across treatment groups?[J]. Stat Med, 2009, 28(9): 1420–1423. <https://doi.org/10.1002/sim.3565>.
- 19 潘雨奇. 基于改进匹配算法的因果效应估计[D]. 武汉: 华中科技大学, 2023. DOI: [10.27157/d.cnki.ghzku.2023.000185](https://doi.org/10.27157/d.cnki.ghzku.2023.000185).
- 20 北京爱琴海乐之技术有限公司. NoteExpress V3.0 功能图解[Z]. 201436.
- 21 Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider[J]. Int J Epidemiol, 2010, 39(2): 417–420. DOI: [10.1093/ije/dyp334](https://doi.org/10.1093/ije/dyp334).
- 22 Glymour MM, Weuve J, Chen JT. Methodological challenges in causal research on racial and ethnic patterns of cognitive trajectories: measurement, selection, and bias[J]. Neuropsychol Rev, 2008, 18(3): 194–213. DOI: [10.1007/s11065-008-9066-x](https://doi.org/10.1007/s11065-008-9066-x).
- 23 VanderWeele TJ, Shpitser I. A new criterion for confounder selection[J]. Biometrics, 2011, 67(4): 1406–1413. DOI: [10.1111/j.1541-0420.2011.01619.x](https://doi.org/10.1111/j.1541-0420.2011.01619.x).
- 24 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](https://doi.org/10.1093/biomet/asx009).
- 25 杨宇琳, 沈瑞博, 李振翔, 等. 多不饱和脂肪酸与结直肠癌的关系: 一项两样本孟德尔随机化研究[J]. 数理医药学杂志, 2024, 37(7): 490–499. [Yang YL, Shen RB, Li ZX, et al. The relationship between polyunsaturated fatty acids and colorectal cancer: a two-sample Mendelian randomization study[J]. Journal of Mathematical Medicine, 2024, 37(7): 490–499.] DOI: [10.12173/j.issn.1004-4337.20240500](https://doi.org/10.12173/j.issn.1004-4337.20240500).
- 26 VanderWeele TJ. Principles of confounder selection[J]. Eur J Epidemiol, 2019, 34(3): 211–219. DOI: [10.1007/s10654-019-00494-6](https://doi.org/10.1007/s10654-019-00494-6).
- 收稿日期: 2025 年 03 月 21 日 修回日期: 2025 年 06 月 03 日
本文编辑: 杨 燕 沈静怡