·综述·

稀有真实世界数据驱动的目标试验模拟在药物再利用中的研究进展

李博生，黄　璇，王文烜，杨文韫，言方荣

中国药科大学理学院生物统计系（南京 210009）

【摘要】针对超说明书用药或共病合并用药的稀有真实世界数据，研究者探索通过构建目标试验模拟以开展针对目标适应证的药物再利用研究。这类研究的成功不仅要求周密的设计，还必须严格按照预设方案与标准化流程执行。在目标试验模拟的设计阶段，核心要素包括明确入排标准、选择试验与对照药物并确定治疗分配时间、选定目标适应证合适的有效性评估指标、明确目标因果效应，以及制定有效的混杂因素校正策略。试验的执行则应遵循一系列严谨步骤：从受试者筛选与受试药物信息提取，到构建试验组与对照组，再到目标试验的模拟操作，最终通过统计推断生成药物再利用的研究假设。在此过程中，倾向性评分计算及其相关的分层、匹配与加权校正技术，对于有效减弱混杂因素对目标因果效应估计的干扰发挥着至关重要的作用。近年来，目标试验模拟的方法学不断创新，特别是在倾向性评分的计算方面。研究者引入了因果学习等先进的机器学习方法进行变量筛选，并积极探索了分类回归树、支持向量机、深度学习等新的数智化技术应用于倾向性评分计算。目前，基于真实世界数据开展的目标试验模拟在药物再利用研究领域已取得显著成果，尤其在心血管疾病、代谢性疾病、阿尔茨海默病、癌症等疾病研究中已展现出广阔的应用前景。

【关键词】目标试验模拟；真实世界数据；倾向性评分；药物再利用；数智化技术

【中图分类号】R 195.1　　【文献标识码】A

Research progress on rare real-world data-driven target trial emulation for drug repurposing

LI Bosheng, HUANG Xuan, WANG Wenxuan, YANG Wenyun, YAN Fangrong

Research Center of Biostatistics and Computational Pharmacy, China Pharmaceutical University, Nanjing 210009, China

Corresponding author: YAN Fangrong, Email: f.r.yan@163.com

【Abstract】For rare real-world data involving off-label drug use or comorbidity-associated polypharmacy, researchers have increasingly adopted target trial emulation to investigate drug repurposing for target indications. The success of such studies hinges on rigorous trial design and strict adherence to predefined protocols and standardized pipelines. Key elements in the trial design include the precise definition of inclusion and exclusion criteria, the selection of trial and control drugs and determination of treatment allocation time, the determination of appropriate efficacy endpoints for the target indication, the identification of causal estimands, and the development of

DOI: 10.12173/j.issn.1005-0698.202501061

基金项目：国家自然科学基金面上项目（82273735）；国家自然科学基金青年基金项目（82304252）；中央高校基本科研业务费专项资金资助（2632023FY05）

通信作者：言方荣，博士，教授，博士研究生导师，Email：f.r.yan@163.com

robust strategies for confounding adjustment. The execution of the trial follows a structured process: screening eligible subjects, extracting relevant drug exposure data, constructing treatment and control groups, emulating the target trial, and ultimately generating hypotheses for drug repurposing through statistical inference. Propensity score methods, including stratification, matching and weighting techniques, are critical tools for addressing confounding bias and ensuring accurate estimation of causal effects. In recent years, creative progress has been made in target trial emulation, particularly in the calculation of propensity scores. Researchers have adopted advanced machine learning techniques, to enhance variable selection and have actively explored the use of innovative methods of digital intelligence technology like classification and regression trees, support vector machines, and deep learning for the application of propensity score calculation. Target trial emulation based on real-world data has achieved remarkable advancements in drug repurposing, demonstrating broad application prospects, particularly in cardiovascular diseases, metabolic disorders, Alzheimer's disease, and cancer.

【Keywords】Target trial emulation; Real-world data; Propensity Score; Drug repurposing; Digital intelligence technology

近年来，药物再利用作为药物研发领域的重要方向，在挖掘已上市药物或研发阶段药物的新治疗用途方面展现出独特价值[1]。与传统全新药物发现相比，该策略既能减少早期开发阶段的资源消耗，又可缩短研发周期、降低开发成本，同时提升监管审批成功率，具有显著优势[2]。值得关注的是，基于队列研究等观察性研究获取的真实世界数据，通过回顾性分析生成真实世界证据，已成为药物再利用假设生成的关键手段[2–3]。然而，基于真实世界数据的药物再利用研究仍面临两大核心挑战，即数据稀疏性问题与观察性研究的因果推断难题。

真实世界数据来源于日常诊疗记录中患者的健康状况及医疗行为信息[4]，主要存储于电子健康记录、医保报销数据库等系统中。值得关注的是，适用于药物再利用研究的真实世界数据往往具有显著稀疏性，其本质源于数据生成场景的临床特殊性：一方面，针对缺乏标准治疗方案的疾病，临床医生可能基于药物作用机制与患者特征实施超说明书用药，但此类有明确治疗意图的超说明书用药行为在真实诊疗中发生率较低，导致用药前后完整的临床状态记录稀缺；另一方面，部分患者因共病治疗需求接受合并用药，若合并用药与目标适应证的病情变化存在潜在关联，其用药记录可能蕴含再利用价值，但此类共病合并用药在临床实践中同样罕见。这一稀有性特征意味着，为确保药物再利用研究的可靠性，研究者需要从大规模真实世界数据中筛选出足够数量且符合入排标准的受试者。

在因果推断层面，利用真实世界数据生成药物再利用假设面临双重障碍：首先，观察性研究因缺乏随机化易受混杂因素干扰，导致对目标因果效应的统计推断存在偏差；其次，研究设计的缺陷（如随访起点设置不当[3, 5]）可能引入人为偏倚，削弱结论的可信度[6]。为应对这些挑战，研究者可采用基于真实世界数据的目标试验模拟方法[6]。该方法通过构建假设性随机对照试验框架，明确定义试验设计要素与执行方案，并基于真实世界数据模拟试验流程；同时结合倾向性评分分层、匹配或加权等方法控制混杂因素，实现对药物目标因果效应的准确估计[6–7]。

本文拟系统解析药物再利用领域目标试验模拟的方法学框架，全面总结该领域最新研究进展（如结合机器学习计算倾向性评分的新方法），并通过典型案例深入剖析其应用价值与共性挑战。最后，着重探讨我国使用目标试验模拟进行药物再利用研究的特殊难题，提出本土化解决方案，并扼要概述突破传统倾向性评分框架的新型目标试验模拟技术。

1 目标试验模拟的方法学框架

为依托稀有的真实世界数据，针对目标适应证开展基于目标试验模拟技术的药物再利用研究，不仅要求在目标试验模拟设计方案中明确界定所有关键要素，还务必严格遵循既定方案，按照规范统一的流程执行试验。在此过程中，科学应用有效的校正方法以控制真实世界数据中混杂因素对目标因果效应估计值的影响，是保证研究

可靠性的关键环节[7]。为助力国内研究者便捷地应用该技术进行研究，本节将系统阐述目标试验模拟的核心设计要素、实施步骤，以及倾向性评分方法在目标因果效应估计中的具体应用机制。

1.1 目标试验模拟的关键设计要素

与人用药品注册技术国际协调会临床试验统计学指导原则增补文件《临床试验中的估计目标与敏感性分析》[8]阐述的临床试验五大要素（人群、治疗策略、终点、群体层面汇总指标、伴发事件及其处理策略）类似，在药物再利用研究的目标试验模拟框架设计中，同样需预先明确五个核心要素：入排标准、治疗分配、有效性结果、目标因果效应、混杂因素及其相应的校正策略。

1.1.1 入排标准

在目标试验模拟框架下的药物再利用研究中，研究设计需依据目标适应证特征与潜在受益人群特点，科学构建入排标准体系，以便从真实世界数据中精准筛选合格受试者。入排标准的制定需综合考量疾病诊断标准、临床分期、共病状态及用药史等关键维度。另外，鉴于目标试验模拟多采用回顾性队列研究数据，受试者筛选过程中对入排标准的应用需严格限定于基线期数据（即受试者接受试验药物或对照药物治疗前的临床观测数据）[3, 5]。

1.1.2 治疗分配

在目标试验模拟设计中，治疗分配的核心在于试验药物与对照药物的选择。试验药物的选择可采用两种策略：直接选用特定超说明书用药[6, 9]，或基于真实世界数据中的合并用药清单按顺序进行挑选[10]。对照药物的选择遵循以下原则：若目标适应证存在标准治疗方案，则优先选用标准治疗药物[6, 9]；若无标准方案，可选择同类药物，如相同解剖学、治疗学及化学分类系统（Anatomical Therapeutic Chemical Classification System，ATC）分类层级的药物，或从真实世界数据中随机选取非试验药物作为对照[3, 5, 10]。为确保研究可靠性，目标试验模拟要求对照组受试者不得同时使用试验药物，并根据试验组与对照组的受试者数量及用药时间，设定是否纳入分析的限制条件[3, 5, 10]。

在目标试验模拟中，治疗分配需明确定义治疗起始时间。该时间点是受试者首次接受试验药物或对照药物的时刻，标志着基线期终止与治疗期启动[3]。类似随机对照试验，目标试验模拟中受试者在治疗起始后可能出现调整初始治疗方案的情况。为与随机对照试验的意向性治疗分析原则保持一致，目标试验模拟中治疗效应的因果推断应当始终基于治疗起始时分配的药物确定受试者分组归属[5]。

1.1.3 有效性结果

在目标试验模拟中，有效性结果是指真实世界数据中能准确反映药物疗效的临床终点指标。有效性结果的选择应与研究目的和目标适应证密切关联，且需经临床研究专家、药物流行病学家和生物统计学家共同讨论后预先确定。与传统临床试验相同，目标试验模拟可采用生存型结果、连续型结果和二分类结果作为有效性结果。

1.1.4 目标因果效应

在目标试验模拟中，目标因果效应是通过比较接受试验药物与对照药物的受试者群体的整体有效性结果来定义的，旨在量化试验药物相对于对照药物的疗效差异。该效应的具体度量方式取决于有效性结果类型：连续型结果采用均值差来表征；二分类结果采用发生率差异来衡量；生存型结果通常采用风险比来评估。

1.1.5 混杂因素及其校正方法

在目标试验模拟中，当基线协变量同时影响治疗分配与有效性结果时，这些变量即构成混杂因素，可能导致试验组与对照组间的目标因果效应估计产生偏倚。因此，为获得对目标因果效应的可靠统计推断，必须通过有效方法校正这些混杂因素。倾向性评分方法（如分层、匹配和加权）是目标试验模拟中最常用的混杂校正方法。关于这些方法的基本原理和适用场景将在本文“1.3.1”一节详细阐述。

完成混杂因素校正后，必须通过组间协变量平衡性诊断来验证校正效果，并确认目标因果效应估计的有效性。具体诊断方法详见“1.3.1”。当混杂因素校正充分且协变量平衡性达标时，即可基于预设有效性结果类型和所选校正方法，使用真实世界数据完成目标因果效应估计。相关估计量的计算方法详见“1.3.2”。

1.2 目标试验模拟的实施步骤

基于“1.1”中概述的关键要素完成目标试验模拟方案设计后，药物再利用研究通常需依次完

成受试者筛选与药物提取、试验组与对照组构建、模拟目标试验，以及统计推断与假设生成这四个核心环节[3, 10]。

1.2.1 受试者筛选与药物提取

根据预设入排标准，首先从真实世界数据库中筛选目标适应证的患者群体。继而基于患者用药记录提取所有单一药物使用数据，同时排除使用人数不足或用药时间过短的用药记录[3, 10]。随后根据治疗起始时间划分基线期与治疗期，并通过核对基线特征与入排标准的一致性，确保最终入选受试者完全符合研究要求。

1.2.2 试验组与对照组构建

依据目标试验模拟方案规定的试验药物与对照药物选择标准及计划模拟次数，编制试验药物与对照药物的配对表。对照组通过从筛选出的受试者中随机抽取未接受相应试验药物治疗的个体构建，同时根据计划模拟次数，在配对表中允许特定试验药物与其对照药物配对的重复出现。

完成配对表制定后，需对每对药物组合在筛选后的受试者真实世界数据库中进行检索，分别构建试验组与对照组的受试者数据子集（涵盖有效性结果与基线期高维混杂因素[3]），为后续目标试验模拟提供数据基础。

1.2.3 模拟目标试验

针对每对提取的试验组与对照组真实世界数据子集，需执行目标试验模拟。为保障目标因果效应统计推断的准确性，需实施有效的混杂因素校正。具体操作中，根据目标试验模拟方案要求，基于数据子集中每位受试者的基线协变量与治疗分配标记计算倾向性评分，并采用适配的校正方法减弱混杂因素对统计推断的潜在影响。校正后需系统评估校正效果，即验证试验组与对照组校正后基线协变量分布是否达到平衡状态[3, 10]。

平衡性诊断达标时方可推进后续目标试验模拟流程；若未达标，则需排除试验组与对照组中倾向性评分极端异常的受试者，仅保留两组评分重叠区间内的个体，随后按方案要求重新进行倾向性评分校正，直至满足平衡性标准[3]。当混杂因素校正完备且通过平衡性诊断后，可基于所选倾向性评分方法构建相应估计量，实现对目标试验模拟方案中预设因果效应的准确估计[3, 10]。

1.2.4 统计推断与假设生成

针对每种待评估的试验药物，首先计算其所有相关目标试验模拟中因果效应估计量的平均值，作为该药物相对于方案选定对照药物的目标因果效应估计值[10]。随后采用 Bootstrap 方法对目标试验模拟的真实世界数据子集进行重抽样，并在每次重抽样后重新执行目标试验模拟与因果效应估计[10]。通过整合不同目标试验模拟重抽样获得的因果效应估计量均值，构建因果效应置信区间并计算 Bootstrap P 值[3, 10]。最终将 Bootstrap P 值与多重性校正后的显著性水平进行对比，判定试验药物与选定对照药物间的疗效差异是否具有统计学意义[10]。若差异有统计学意义，则可生成药物再利用假设，提示试验药物在目标适应证中可能具备显著的治疗效果。

通过上述流程，目标试验模拟能有效利用电子健康记录、政府医保数据库等来源的非适应证用药真实世界数据，系统性评估老药在潜在新适应证中的疗效，为药物再利用研究提供可靠证据支持。

1.3 倾向性评分运用机制

在目标试验模拟中，倾向性评分通常与适配方法联用，通过校正可观测混杂因素降低其对目标因果效应估计的干扰。倾向性评分定义为给定基线协变量 X 条件下受试者分配至特定组别的概率，其中 $P(Z=1|X)$ 代表受试者分配至试验组的条件概率，$P(Z=0|X)$ 则对应受试者分配至对照组的条件概率。

实际操作中，倾向性评分通常通过基线协变量与组别标记间的 Logistic 回归模型进行估计。鉴于目标试验模拟常涉及高维协变量引发的多重共线性问题，近期研究引入变量选择方法筛选高维混杂因素中的核心变量以优化评分估计[3, 10–11]。此外，机器学习及深度学习技术也逐渐应用于个体倾向性评分估计。相关研究进展详见本文“2”章。

完成倾向性评分估计后，可采用分层、匹配或加权等方法校正已知混杂因素，校正后需系统评估协变量平衡性。本节将概述常规校正方法与平衡性诊断标准。完成校正并通过平衡性诊断后，即可基于校正方法构建对应估计量，对方案定义的目标因果效应进行参数估计。本节亦将解析目标试验模拟中常用估计量的计算方法。

1.3.1 混杂因素校正暨平衡性诊断

（1）倾向性评分分层：在对目标试验模拟

中受试者的倾向性评分完成估计后，首先将试验组与对照组的评分汇总并按升序排列，然后均等划分为若干区间（常规建议至少 5 个区间[12-13]，若组间协变量无系统性差异则可终止细分[14]）。确定区间划分后，受试者依评分归入对应分层，后续分析参照分层随机对照试验方法执行。该方法的有效性源于其通过“局部随机化”近似实现组间可比性，显著降低混杂偏倚并提升因果效应估计稳健性。其操作灵活性体现在通过分层数量调整适配数据结构特征，并借助平衡性诊断工具快速识别协变量残留的不平衡，从而为后续策略优化提供依据。然而，其效能受限于分层数量（过少致残余混杂，过多致样本稀疏）及数据结构特性（极端评分或高维混杂会削弱校正精度），故更适用于大样本数据的粗粒度平衡预处理，后续精细分析需结合倾向性评分加权、匹配等方法加以实现。

完成倾向性评分分层后，需系统评估试验组与对照组基线协变量平衡性[15]。本文推荐采用 Rosenbaum 等[12]提出的经典诊断法：对每个协变量构建包含倾向性评分分层、治疗指示变量及两者交互项的两因素方差分析模型。若治疗变量或交互项对协变量的影响具有统计显著性，则表明至少存在一个分层未达平衡；反之则通过平衡性诊断[15]。

（2）倾向性评分匹配：在对目标试验模拟中受试者的倾向性评分完成估计后，可依据预设的试验组与对照组样本分配比例，筛选倾向性评分相近的受试者进行匹配，从而构建严格匹配的样本集，为后续统计推断奠定坚实基础。关于具体操作流程，Austin[16]提出的匹配策略包含以下核心步骤：对倾向性评分进行 logit 转换；基于评分对数优势的 20% 定义匹配容差范围，筛选组间差异落在该区间的受试者进行配对；同时可采用贪婪最近邻匹配算法优化匹配效率。

倾向性评分匹配通过构建“伪随机”样本降低组间协变量差异，有效控制混杂偏倚，且可直接沿用随机对照试验分析方法提升因果效应估计可信度。该方法的灵活性体现在可通过调整匹配容差范围与匹配机制适配不同数据结构。但需注意，当试验组与对照组倾向性评分重叠度较低时，该方法易导致大量样本丢失而降低统计效能，且大样本下计算复杂度较高。因此更适用于小样本场景以最大化保留具有可比性的个体，或作为敏感性分析工具，与加权、分层等方法结合验证结论稳健性。

经过倾向性评分匹配后，需通过标准化差异评估组间协变量平衡性[16]。针对连续型协变量，标准化差异计算公式[16]为：

$$d=\frac{\bar{x}_{\mathrm{T}}-\bar{x}_{\mathrm{C}}}{\sqrt{\frac{s_{\mathrm{T}}^2+s_{\mathrm{C}}^2}{2}}} \quad \text{公式（1）}$$

其中，$\bar{x}_T$ 与 $\bar{x}_C$ 分别为试验组与对照组的协变量样本均值，而 s_{T}^2 与 s_{C}^2 为对应的样本方差。二分类协变量的标准化差异按公式 2 计算：

$$d=\frac{\left(\hat{p}_{\mathrm{T}}-\hat{p}_{\mathrm{C}}\right)}{\sqrt{\frac{\hat{p}_{\mathrm{T}}\left(1-\hat{p}_{\mathrm{T}}\right)+\hat{p}_{\mathrm{C}}\left(1-\hat{p}_{\mathrm{C}}\right)}{2}}} \quad \text{公式（2）}$$

其中，$\hat{p}_C$ 和 $\hat{p}_T$ 分别为两组中二分类协变量事件的发生率估计。当所有协变量标准化差异均低于 0.1 时[17]，可认定通过平衡性诊断。

（3）倾向性评分加权：在准确估计目标试验模拟中每位受试者的倾向性评分后，可采用逆概率加权（或其改进版本）对真实世界数据子集中的观测进行加权处理，构建基线协变量分布更均衡的伪随机试验[10]，为后续目标因果效应的统计推断奠定坚实基础。逆概率加权的常规表达形式[14]如下：

$$W=\frac{Z}{P\left(Z=1\middle|\boldsymbol{X}\right)}+\frac{1-Z}{1-P\left(Z=1\middle|\boldsymbol{X}\right)} \quad \text{公式（3）}$$

其中，符号 Z 和 X 依旧分别代表受试者的治疗指示变量与基线协变量。实际应用中为提高稳健性，常采用其改进形式——稳定化逆概率加权法[10-11]：

$$W=\frac{Z\times P(Z=1)}{P(Z=1|\boldsymbol{X})}+\frac{(1-Z)\times P(Z=0)}{1-P(Z=1|\boldsymbol{X})} \quad \text{公式（4）}$$

倾向性评分加权法通过为受试者赋予特定权重构建协变量分布均衡的“伪总体”，其核心优势在于全局调整组间协变量差异时无需剔除样本，这一特性使其尤其适用于匹配法因样本不重叠失效或分层法导致样本稀疏的场景。在此基础上，通过稳定化权重降低倾向性评分接近 0 或 1 时出现的极端值引发的方差膨胀，该方法得以在保留所有个体信息的基础上协调偏倚与方差的矛盾。更为重要的是，相较于分层法和匹配法，其优势还突出表现为对高维协变量的强适应性，以

及与结果模型结合形成双重稳健估计的能力，最终系统性提升因果效应推断的可靠性。但需强调的是，残余极端权重可能导致估计不稳定，需依赖截断或平滑技术加以控制。

在经过倾向性评分加权处理后，试验组与对照组间各协变量的平衡状态仍通过标准化差异进行诊断，但需将传统标准化差异计算中的估计量替换为相应的加权版本[10–11]。具体而言，对于连续型基线协变量，试验组与对照组各自的协变量加权样本均值及方差可按以下公式[10–11]计算：

$$\bar{x}_T=\frac{\sum_{i=1}^{n}w_i x_i z_i}{\sum_{i=1}^{n}w_i z_i}\text{ 且 }\bar{x}_C=\frac{\sum_{i=1}^{n}w_i x_i (1-z_i)}{\sum_{i=1}^{n}w_i (1-z_i)} \quad \text{公式（5）}$$

$$s_T^2=\frac{\sum_{i=1}^{n}w_i z_i}{\left(\sum_{i=1}^{n}w_i z_i\right)^2-\sum_{i=1}^{n}z_i w_i^2}\sum_{i=1}^{n}z_i w_i\left(x_i-\bar{x}_T\right)^2$$

$$s_C^2=\frac{\sum_{i=1}^{n}w_i (1-z_i)}{\left[\sum_{i=1}^{n}w_i (1-z_i)\right]^2-\sum_{i=1}^{n}(1-z_i) w_i^2}\sum_{i=1}^{n}(1-z_i) w_i\left(x_i-\bar{x}_C\right)^2 \quad \text{公式（6）}$$

其中，n 代表目标试验模拟中受试者总数，w_i、x_i 及 z_i 则分别对应第 i 位受试者的逆概率加权（或其改进版本）、连续型基线协变量取值以及治疗指示变量（z_i=1 为试验组，z_i=0 为对照组）。对于二分类基线协变量，试验组与对照组各自的协变量加权发生率估计量应按公式 7[11]计算：

$$\hat{p}_T=\frac{\sum_{i=1}^{n}w_i x_i' z_i}{\sum_{i=1}^{n}w_i z_i}\text{ 且 }\hat{p}_C=\frac{\sum_{i=1}^{n}w_i x_i' (1-z_i)}{\sum_{i=1}^{n}w_i (1-z_i)} \quad \text{公式（7）}$$

其中，x_i' 表示目标试验模拟中第 i 位受试者的二分类变量，取值为 0 或 1。

1.3.2 常见因果效应估计量

目标试验模拟中目标因果效应的估计量需根据有效性结果的类型（连续型、二分类或生存型）及所采用的倾向性评分方法（分层、匹配或加权）进行适配。本节针对上述 3 类有效性结果，分别阐述不同倾向性评分方法对应的估计量形式。

当有效性结果为连续型或二分类变量时，目标因果效应通常定义为试验组与对照组间的均值差异或发生率差异。若采用倾向性评分分层法校正混杂因素，连续型结果的因果效应估计量为各层内试验组与对照组均值差异的加权平均，二分类结果则为各层发生率差异的加权平均，权重均为各层样本占比[14, 18]。当使用倾向性评分匹配法时，连续型结果的估计量是匹配样本的组间均值差异，二分类结果的估计量是匹配样本的组间发生率差异[14, 18]。若采用倾向性评分加权法，连续型结果的估计量为试验组与对照组加权均值差异（计算参考公式 5），二分类结果为加权发生率差异（计算参考公式 7）[14, 19]。

当有效性结果为生存数据时，目标因果效应通常定义为试验组与对照组间的生存风险比。当采用倾向性评分分层法时，可通过构建以倾向性评分分层为分层变量的分层 Cox 比例风险模型，并将治疗指示变量作为回归项，所得风险比即为目标因果效应估计量[20]。当使用倾向性评分匹配法时，需基于匹配样本构建 Cox 比例风险模型并纳入治疗指示变量，其风险比可作为目标因果效应的估计量[20]。若应用倾向性评分加权法，则需通过逆概率加权或其改进版本构建加权 Cox 模型，在部分似然函数中对乘积项施加权重（即取对应指数幂），并纳入治疗指示变量作为回归项，所得风险比即为目标因果效应估计量[20–21]。

2 倾向性评分计算的新方法

近年来，药物再利用领域的目标试验模拟方法学研究进展主要体现在基于机器学习技术的倾向性评分计算新方法，其技术路径可归纳为两大方向。第一类方法聚焦于应用机器学习技术进行精确的变量选择，通过从高维基线协变量中识别对治疗 – 有效性结果间因果效应具有显著混杂作用的关键变量，继而基于筛选出的主要混杂变量及治疗指示变量构建 Logistic 回归模型，实现倾向性评分的精确估计。此类方法通过减弱无关协变量的干扰，有效缓解了传统建模中多重共线性偏倚问题，显著提升了评分估计的稳健性[10]。第二类方法则突破传统建模范式，将倾向性评分估计重构为分类问题，直接以基线协变量与治疗指示变量作为输入特征，采用多种机器学习算法进行端到端的倾向性评分预测。该方法摒弃了传统 Logistic 回归模型中倾向性评分对数优势与基线协变量间的线性约束，能够更灵活地刻画倾向性评分与基线协变量间复杂的非线性关联，从而在复杂场景下实现更高精度的倾向性评分估计。本节将系统梳理这两类倾向性评分计算新方法的研究进展。

2.1 基于机器学习的变量筛选方法

近年来，目标试验模拟研究中逐渐发展出结合领域知识与数据驱动因果学习算法的倾向性评分协变量选择策略，该策略已成为基线协变量筛选中较为流行的方案[10]。具体而言，首先基于专家领域知识从数据库中筛选出可能与治疗分配或有效性结果相关的基线协变量，并初步构建包含3类关键变量的因果图：混杂变量（同时影响治疗分配与有效性结果，需纳入倾向性评分模型调整）、中介变量（位于治疗分配与结果的因果路径中，调整将阻断因果效应）、碰撞变量（治疗与结果的共同后果，调整可能引入选择偏倚）。倾向性评分建模时需排除中介变量与碰撞变量，仅保留潜在混杂变量[10]。鉴于领域知识可能存在缺失或误判，通常需联合应用基于数据驱动的约束因果结构学习算法（如稳健PC算法）对因果图进行修正，最终依据修正后的因果图确定需调整的混杂变量集合[10]。该策略通过自动化筛选高维变量，有效缓解了真实世界数据（如含数百个指标的电子健康记录）中的维度灾难问题，同时维持临床可解释性。针对作用机制尚未明确的适应证开展探索性研究时，结合领域知识与因果学习的变量筛选技术可深度挖掘潜在混杂因素，弥补先验知识不足的缺陷。

另外，在倾向性评分建模实践中，更为精细的模型选择（涵盖变量选择）与优化技术逐渐受到重视。以新型“机器学习倾向性评分模型训练与选择的交叉验证算法”[10]为例，该方法在训练阶段将最小绝对收缩与选择算子（least absolute shrinkage and selection operator，LASSO）和岭回归的正则化超参数作为关键调优对象，并在交叉验证中实施双重评估：基于验证集曲线下面积（area under the curve，AUC）指标评估预测精度，同时通过协变量标准化差异约束组间平衡性。LASSO正则化通过稀疏化变量选择降低维度，岭回归正则化则抑制多重共线性引起的参数估计偏倚，两者协同增强模型泛化能力[10]。该双重评估策略即使应用于正则化Logistic回归等简约模型，亦可使倾向性评分估计精度达到与梯度提升、深度神经网络等复杂模型相当的水平[10]。这挑战了模型复杂度与性能正相关的传统认知，凸显了正则化技术与双重评估策略协同使用在变量选择与模型优化中的核心价值[10]。然而，需要强调的是，该算法的优势依赖于存在强预测变量及协变量高相关性的数据结构假设。当协变量对治疗分配呈弱独立影响，或存在显著非线性关联与交互效应时，其可能难以捕获协变量与倾向性评分间的深层关联。当前研究正尝试融合领域知识引导的变量筛选与机器学习非线性建模以突破此局限[22]。

2.2 基于机器学习的倾向性评分计算方法

在倾向性评分估计中，传统的Logistic回归模型虽被广泛应用，但其局限性也逐渐显现。为了克服这些限制，机器学习领域的多种方法被发掘并尝试应用于此领域。这些方法的选择与应用需紧密贴合数据特征，以确保估计结果的准确性和可靠性。

Logistic回归作为倾向性评分估计的经典方法，在医疗研究中占据重要地位，尤其适用于低维数据与机制明确的流行病学队列分析。它以其计算简洁性和对线性关系的良好处理能力而广受欢迎，并且模型具有较强的可解释性。然而，Logistic回归在估计倾向得分时，要求基线协变量与对数优势比之间存在严格的线性关系，并且需要确保不遗漏重要的交互作用项，以避免模型误设可能导致的偏倚[23]。此外，Logistic回归在处理高维数据、复杂的非线性关系以及潜在的交互作用时可能显得力不从心，这限制了其在某些现代医疗研究中的应用范围[23]。

支持向量机（support vector machine，SVM）通过核技巧将原始协变量映射至高维特征空间，并构建最大间隔超平面进行分类决策。通过径向基核等非线性核函数，SVM能够解析协变量与处理分配间的复杂关联，在处理倾向性评分与基线协变量间的复杂非线性关系及不可加性问题时具有显著优势。其无需预设协变量与组别间的参数化函数关系，为复杂数据分析提供了灵活性。并且，通过正则化项控制模型复杂度则赋予SVM处理高维数据的适应性与稳健性[23]。然而，传统SVM主要面向模式分类任务，虽能判定样本组别归属，却无法直接输出倾向性评分。为此改进的导入向量机（import vector machine，IVM）通过贝叶斯框架校准SVM输出，从而直接输出符合概率公理化定义的倾向性评分估计。另一方面，核函数选择仍是SVM应用的关键挑战，需依据数据特征手动设定以实现超平面划分和协变量有效转换。因此，SVM更适用于具有强非线性关联的中

等规模因果推断任务。与随机森林等特征选择算法结合可增强核函数构建的生物学合理性及模型解释性，进而提升因果推断的准确度与可靠性[23]。

分类回归树（classification and regression trees，CART）及其集成方法在倾向性评分估计中得到了广泛应用。CART 通过递归分区直接构建分类规则，实现对协变量空间的非参数划分，进而叶节点样本比例可直接转化为倾向性评分估计，这使得 CART 在临床应用中具备较高的可解释性和实施便捷性[23]。然而，单一 CART 模型易因过度分割产生过拟合，故对训练数据噪声敏感[24]。为解决这一问题，研究者们提出了多种改进方法。Pruned CART 通过控制树节点数量来防止过拟合，从而提升模型的泛化能力[24]。Bagged CART 采用自助抽样技术构建多棵 CART，并聚合其预测结果，显著降低倾向性评分估计的偏倚，尤其适用于处理存在测量误差的医疗数据[23-24]。随机森林在 bagged CART 的基础上进一步引入特征随机子集选择，降低树间相关性，不仅在高维数据中优异地平衡组间协变量，还能自动捕捉协变量间的交互效应，减轻了在参数模型中手工设定交互项的负担[23-24]。此外，提升学习算法（如 boosted CART）通过迭代加权组合多个弱分类器构建强分类器，展现出更强的抗过拟合能力，在数据存在复杂非线性关联时能显著降低因果效应估计的偏差，甚至还能组合多个不直接输出类别概率的弱分类器（如决策树、SVM）并转化为能输出类别概率（即估计倾向性评分）的强分类器[23-24]。但是，应用提升学习算法时也需要谨慎设置学习率并采用早停法以防止过拟合[23-24]。

神经网络凭借其多层感知机架构与非线性激活函数的精妙结合，在估计倾向性评分领域展现出了突出的优势。在高维数据情景下，即便单个基线协变量对治疗分配概率的影响微乎其微，神经网络依然能够凭借其复杂的隐藏层网络结构，敏锐捕捉这些细微信号，并实现全局信息的深度整合，从而显著提升预测模型的性能[23]。更进一步，神经网络的通用近似定理为其提供了坚实的理论基础，确保了其对任意光滑函数的卓越逼近能力，这有效规避了传统方法中因手动设定多项式阶数或交互项而可能引发的模型误设风险[23]。这些特性使神经网络成为复杂高维数据分析中估计倾向性评分的理想选择。尽管早期神经网络在训练与优化方面面临诸多挑战，但随着近年来人工智能技术与微电子技术的飞速发展，这些难题已得到了有效克服。

Liu 等[25]针对目标试验模拟中受试者基线期长、协变量随时间动态变化的问题，设计了一种融合长短期记忆神经网络（long short-term memory，LSTM）与注意力机制的深度学习模型，显著提升了倾向性评分的估计精度。实验结果显示，当通过逆概率加权校正混杂因素时，相比传统 Logistic 回归模型，上述模型在偏差校正及治疗效应估计的准确性方面具有显著优势，同时保留了识别关键混杂因素的可解释性[25]。然而，在短期基线静态数据或显性线性机制场景下，Logistic 回归仍因其计算效率和参数简洁性而具有一定竞争力，与合适的变量选择方法结合也能将目标因果效应估计偏倚控制在一定范围并防范过拟合风险[10]。

Ghosh 等[26]结合稀疏自编码器的降维能力和深度学习的拟合优势，提出了一种针对高维医学数据的因果推断框架——稀疏自编码器增强的深度倾向网络（deep propensity network using a sparse autoencoder，DPN-SA）。该框架通过 3 个关键步骤实现了高效因果推断：首先利用稀疏自编码器对原始高维协变量进行非线性降维，旨在保留关键混杂因素的同时显著降低数据维度；随后将降维后的特征输入深度神经网络，通过稳定训练过程，输出精确的个体化倾向评分；最后采用 Adam 优化器进行端到端的联合优化，旨在平衡降维保真度与因果效应估计的精度[26]。DPN-SA 能有效应对高维协变量、非线性或非平行的治疗分配，以及残余混杂等挑战，从而提升了倾向性评分与治疗效应的估计精度[26]。实验证明，相较于传统的 Logistic 回归和 LASSO 方法，DPN-SA 在多个数据集上展现了更优的倾向评分估计精度与治疗效果评估稳健性[26]。

Weberpals 等[7]的研究则采用自编码器对高维协变量进行降维，并直接基于自编码器训练产生的嵌入变量进行 Logistic 回归以估计倾向性评分。然而研究发现，这一新策略在性能上并不优于传统的 LASSO 方法。因此，尽管深度学习模型在复杂数据分析中展现出巨大潜力，但在某些特定场景下，传统统计方法仍可能保持竞争力和实用性。

3 案例分析

近年来，目标试验模拟技术已成为推动药物再利用研究的重要工具，针对心血管疾病、代谢性疾病、阿尔茨海默病、癌症等多种目标适应证的应用案例层出不穷。为便于国内同行借鉴与参考，本文对近年来经典的应用案例进行了系统梳理与总结，进而深入分析了目标试验模拟技术在各案例中的应用效果，最后归纳了该技术在实际应用中面临的共同挑战。

3.1 心血管疾病中的应用

2021 年，Liu 等[25]基于目标试验模拟框架，利用 2012—2017 年 MarketScan 医疗保险索赔数据库中的 1 178 997 例冠心病患者的真实世界数据开展了一项创新性的药物再利用研究。该研究通过 LSTM 和逆概率加权法控制混杂因素，成功模拟了随机对照试验，识别出美托洛尔、非洛贝特等 6 种药物以及美托洛尔与氯吡格雷联用等 7 种药物组合可显著改善冠心病患者预后，其中美托洛尔已被证实可通过降低心力衰竭风险改善冠心病患者的预后[25, 27]，非洛贝特则显示出降低糖尿病患者冠心病风险的潜力[25, 28]。

3.2 代谢性疾病中的应用

2022 年，Charpignon 等[29]基于目标试验模拟框架，整合美国研究患者数据库（Research Patient Data Registry，RPDR）与英国临床实践研究数据库（Clinical Practice Research Datalink，CPRD）两大电子健康数据库的真实世界数据，系统比较了 2 型糖尿病患者使用二甲双胍与磺脲类药物的长期疗效差异。研究通过因果推断方法控制混杂因素，并校正死亡对痴呆事件的竞争风险影响，结果显示二甲双胍使用者的全因死亡率与痴呆风险均显著低于磺脲类药物组，且年轻患者（70 岁以下）的认知保护效应更为明显。进一步实验[29]表明，二甲双胍可抑制人神经细胞中与阿尔茨海默病相关的关键蛋白（如骨桥蛋白 SPP1）的表达，而磺脲类药物未观察到类似作用，提示其可能通过非降糖途径（如调节神经炎症或淀粉样代谢）延缓痴呆进程，但临床转化需结合生物标志物验证。

3.3 阿尔茨海默病中的应用

2021 年，Fang 等[30]基于 2012—2017 年 MarketScan 商业医疗保险索赔数据库中 723 万患者的真实世界数据，评估西地那非在阿尔茨海默病中的潜在预防作用。该研究通过倾向性评分分层分析控制年龄、性别、种族及共病等混杂因素，对比西地那非用药组与多个对照药物组（包括地尔硫䓬、氯沙坦、格列美脲及二甲双胍用药人群）的阿尔茨海默病发病风险。结果显示，使用西地那非与阿尔茨海默病发病风险降低存在统计学上的显著相关性[30]。

2023 年，Zang 等[10]整合美国 OneFlorida 电子健康记录与 MarketScan 商业保险数据库，构建包含 1.7 亿患者记录的大规模观察性队列，进一步筛选约 50 万例轻度认知障碍患者，采用高通量目标试验模拟框架系统评估 4 300 余种药物与阿尔茨海默病发病风险的关联。通过机器学习优化的倾向性评分模型选择策略结合逆概率加权方法控制混杂因素，结果发现患者使用泮托拉唑、加巴喷丁、阿托伐他汀、氟替卡松和奥美拉唑与阿尔茨海默病发病风险降低显著相关，提示这些药物在疾病预防中的潜在价值[10]。

2024 年，Yan 等[31]基于目标试验模拟框架，创新性运用 ChatGPT 作为智能假设生成工具，系统筛选阿尔茨海默病药物再利用候选化合物。该研究通过迭代式提示工程获得候选药物后，选取重复频次最高的前 10 种化合物，利用 Vanderbilt 大学医学中心和美国全民健康研究计划两大真实世界临床数据库，采用倾向性评分匹配的回顾性队列研究设计进行验证，多因素 Cox 回归分析结果表明，长期服用二甲双胍、辛伐他汀和洛沙坦与 65 岁以上老年人群阿尔茨海默病发病风险降低显著相关，这一发现为相关已上市药物的疾病预防应用拓展提供了新的循证证据[31]。

3.4 癌症中的应用

2019 年，Wu 等[32]基于目标试验模拟框架，系统分析 1995—2010 年范德比尔特大学医学中心和梅奥诊所的 143 310 例及 98 366 例癌症患者的电子健康记录，评估 146 种非抗癌药物对肿瘤预后的潜在影响。该研究采用两阶段验证设计：首先在范德比尔特队列中评估 146 种长期使用的非抗癌药物，通过多变量 Cox 回归模型（校正人口学、肿瘤分期及合并症等混杂因素）筛选出 22 种与生存期延长显著相关的药物，涵盖他汀类、质子泵抑制剂等六大类别；随后在梅奥诊所独立队列中成功验证其中 9 种药物的保护效应，包括瑞舒伐他汀、奥美拉唑等典型药物。研究[32]进一

步通过系统文献分析证实，这些跨机构验证的药物均存在已知或潜在的抗肿瘤作用机制。

2020 年，Dickerman 等 [33] 基于CPRD 1998—2016 年的数据，建立包含 22 163 例结直肠癌病例及 88 652 例匹配对照的观察性研究队列，系统探讨他汀类药物暴露与疾病风险的关联性。研究采用目标试验模拟框架，通过精确界定治疗启动时间窗与随访周期模拟随机对照试验设计，并运用逆概率加权法校正人口学特征、共病状态及伴随用药等混杂因素，辅以多维度敏感性分析确保结果可靠性。在控制时间依赖性偏倚的基础上，研究发现长期使用他汀类药物可能与结直肠癌风险降低存在关联，且该关联在老年人群及不同用药亚组中呈现一致性趋势 [33]。

3.5 实际应用的挑战

参考上述典型案例，采用目标试验模拟技术开展药物再利用研究时，研究者往往需要应对数据规模需求与数据稀疏性之间的突出矛盾。具体而言，为确保研究结果的可靠性，通常需要评估数百至上千种候选药物，而每种药物又需包含至少数百例用药者的完整数据记录，这就要求从真实世界数据库中筛选数十万例符合入排标准的受试者。然而，药物再利用研究所依赖的超说明书用药和共病合并用药数据在临床实践中本就较为罕见，使得合格样本的获取面临显著困难。为满足研究需求，研究者不得不从区域电子健康档案或医保数据库中提取数以千万计的原始数据记录以供筛选，这就不可避免地大幅提升了数据采集、存储、清洗和检索等环节的技术难度与资源消耗。综上所述，数据规模需求与固有稀疏性矛盾所驱动的高强度数据处理需求，构成了目标试验模拟技术在药物再利用研究中面临的根本性技术障碍。

4 结语

目标试验模拟在药物再利用研究中的应用可为“老药新用”的疗效验证提供初步证据，此类证据或有助于药品监管部门豁免部分临床前有效性研究。此外，“老药”的长期临床应用数据已积累充分的安全性证据，可能为豁免临床前安全性研究及 I 期临床试验剂量探索提供依据。该方法的应用既能加速药物研发进程，亦可显著降低研发成本。为促进我国研究者更高效地利用目标试验模拟开展药物再利用研究，本文系统梳理了其方法学框架（包括关键设计要素、实施流程及倾向性评分应用机制）、新型倾向性评分计算方法，以及该技术成功应用于药物再利用研究的典型案例。然而，当前我国基于目标试验模拟的药物再利用研究仍面临诸多挑战，其中部分问题超越了真实世界大规模稀疏数据治理的常规范畴。

真实世界数据中未知混杂因素的存在可能影响目标试验模拟所得药物再利用结论的可靠性[34]。现有技术仅能用于生成药物再利用假设，而无法达到部分文献提出的“验证假设”程度 [2]。因此，基于目标试验模拟筛选出的潜在适应证药物，仍需遵循现行监管要求开展确证性 III 期临床试验。对于尚处于专利保护期的药物，原研药企可能具有通过目标试验模拟探索新适应证并推动确证性临床试验的动机；然而，针对已上市多年的老药，即使通过目标试验模拟发现其潜在新适应证，常因缺乏商业利益驱动而难以吸引药企资助后续确证性研究。典型案例为“百年老药”阿司匹林虽被证实具有预防结肠癌的潜在疗效，却因商业价值与监管支持不足未能获批肿瘤学适应证 [2]。

为应对未知混杂因素对因果效应推断的干扰，需在目标试验模拟的设计与分析阶段针对性地开发新型因果推断方法。与此同时，各国监管部门亟需探索针对专利过期老药新适应证开发的激励性监管政策，以突破当前“无利可图”导致的研发瓶颈。

另外，我国应用目标试验模拟开展药物再利用的核心挑战集中于真实世界数据的质量与共享机制 [35]。当前国内真实世界数据普遍存在患者预后纵向随访记录缺失、元数据信息透明度不足等问题 [35]。在数据获取层面，电子健康记录、医保数据等关键数据源的开放程度有限，且未建立规范化的公开获取渠道 [35]。目前仅海南真实世界数据研究院（https://hnrws.cn/2750/2811.html）作为国家级平台提供数据协调服务，但具体协作机制尚不明晰。数据获取对非公开渠道的依赖导致研究开展严重受限于研究者的资源网络。此外，我国对真实世界数据使用的审批流程及隐私保护缺乏专门立法，致使研究常面临复杂的伦理审查程序 [35]。更为重要的是，多数研究还需通过国家人类遗传资源管理办公室审批，进一步增加了研究

启动的行政负担[35]。

针对上述问题，Xie 等[35]提出三方面改进建议：通过立法明确数据所有权与审批流程，建立数据安全标准；系统性提升数据质量；构建符合我国医疗体系特点的数据基础设施以保障数据采集的效率与质量。这些建议为破解当前困境提供了重要参考方向。

最后，值得注意的是，尽管本文所讨论的目标试验模拟方法大多基于倾向性评分，但是一些新型深度学习模型在不依赖倾向性评分的情况下也能准确估计目标因果效应。例如，Zhang 等[36]提出的 TransTEE 模型通过 Transformer 架构对协变量和治疗进行编码，并借助交叉注意机制有效调整混杂偏倚，能够高效估计治疗效应。该模型在处理复杂的高维数据和稀疏的临床数据时表现尤为出色，通过自注意力机制捕捉变量间的复杂关系，从而在大规模数据分析中提供准确的治疗效果估计[36]。同样，Liu 等[37]提出的 CURE 框架也利用 Transformer 模型在大规模未标记患者数据上进行预训练，学习丰富的上下文信息，并在标记数据上微调以优化治疗效果的估计。CURE 框架能够自动捕捉复杂的非线性关系，尤其适用于处理高维和稀疏的临床数据，并展现了强大的适应性和估计精度[37]。这些方法为观察性研究中的因果效应估计提供了新的视角和工具，具有广阔的应用前景。

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 Zong N, Wen A, Moon S, et al. Computational drug repurposing based on electronic health records: a scoping review[J]. NPJ Digit Med, 2022, 5(1): 77. DOI: 10.1038/s41746-022-00617-6.

2 Tan GSQ, Sloan EK, Lambert P, et al. Drug repurposing using real-world data[J]. Drug Discov Today, 2023, 28(1): 103422. DOI: 10.1016/j.drudis.2022.103422.

3 Ozery-Flato M, Goldschmidt Y, Shaham O, et al. Framework for identifying drug repurposing candidates from observational healthcare data[J]. JAMIA Open, 2020, 3(4): 536-544. DOI: 10.1093/jamiaopen/ooaa048.

4 国家药品监督管理局 . 真实世界证据支持药物研发与审评的指导原则（试行）[S/OL]. (2020-01-07) [2025-07-23]. https://www.cde.org.cn/zdyz/domesticinfopage?zdyzIdCODE=db4376287cb678882a3f6c89.

5 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available[J]. Am J Epidemiol, 2016, 183(8): 758-764. DOI: 10.1093/aje/kwv254.

6 Matthews AA, Danaei G, Islam N, et al. Target trial emulation: applying principles of randomised trials to observational studies[J]. Br Med J, 2022, 378: e071108. DOI: 10.1136/bmj-2022-071108.

7 Weberpals J, Becker T, Davies J, et al. Deep learning-based propensity scores for confounding control in comparative effectiveness research a large-scale, real-world data study[J]. Epidemiology, 2021, 32(3): 378-388. DOI: 10.1097/ede.0000000000001338.

8 International Council for Harmonisation. Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials E9(R1) [R/OL]. (2019-11-20) [2025-07-23]. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.

9 Urner M, Barnett AG, Li Bassi G, et al. Venovenous extracorporeal membrane oxygenation in patients with acute covid-19 associated respiratory failure: comparative effectiveness study[J]. Br Med J, 2022, 377: e068723. DOI: 10.1136/bmj-2021-068723.

10 Zang C, Zhang H, Xu J, et al. High-throughput target trial emulation for Alzheimer's disease drug repurposing with real-world data[J]. Nat Commun, 2023, 14(1): 8180. DOI: 10.1038/s41467-023-43929-1.

11 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.

12 Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score[J]. J Am Stat Assoc, 1984, 79(387): 516-524. DOI: 10.2307/2288398.

13 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.

14 Imbens GW, Rubin DB, eds. Causal inference in statistics, social, and biomedical sciences[M]. Cambridge: Cambridge University Press, 2015: 274-276, 382.

15 Austin PC. Goodness-of-fit diagnostics for the propensity score model when estimating treatment effects using covariate adjustment with the propensity score[J]. Pharmacoepidemiol Drug Saf, 2008, 17(12): 1202-1217. DOI: 10.1002/pds.1673.

16 Austin PC. 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.

17 Normand SLT, Landrum MB, Guadagnoli E, et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores[J]. J Clin Epidemiol, 2001, 54(4): 387-398. DOI: 10.1016/S0895-4356(00)00321-8.

18 Austin PC, Mamdani MM. A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI

statin use[J]. Stat Med, 2006, 25(12): 2084–2106. DOI: 10.1002/sim.2328.

19 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.

20 Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios[J]. Stat Med, 2013, 32(16): 2837–2849. DOI: 10.1002/sim.5705.

21 Buchanan AL, Hudgens MG, Cole SR, et al. Worth the weight: using inverse probability weighted Cox models in AIDS research[J]. AIDS Res Hum Retroviruses, 2014, 30(12): 1170–1177. DOI: 10.1089/aid.2014.0037.

22 Karpatne A, Jia X, Kumar V. Knowledge–guided machine learning: current trends and future prospects[EB/OL]. (2024–05–01) [2025–07–23]. https://doi.org/10.48550/arXiv.2403.15989.

23 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.

24 Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning[J]. Stat Med, 2010, 29(3): 337–346. DOI: 10.1002/sim.3782.

25 Liu R, Wei L, Zhang P. A deep learning framework for drug repurposing via emulating clinical trials on real–world patient data[J]. Nat Mach Intell, 2021, 3(1): 68–75. DOI: 10.1038/s42256–020–00276–w.

26 Ghosh S, Bian J, Guo Y, et al. Deep propensity network using a sparse autoencoder for estimation of treatment effects[J]. J Am Med Inform Assoc, 2021, 28(6): 1197–1206. DOI: 10.1093/jamia/ocaa346.

27 Fisher ML, Gottlieb SS, Plotnick GD, et al. Beneficial effects of metoprolol in heart failure associated with coronary artery disease: a randomized trial[J]. J Am Coll Cardiol, 1994, 23(4): 943–950. DOI: 10.1016/0735–1097(94)90641–6.

28 Wong TY, Simo R, Mitchell P. Fenofibrate – a potential systemic treatment for diabetic retinopathy?[J]. Am J Ophthalmol, 2012, 154(1): 6–12. DOI: 10.1016/j.ajo.2012.03.013.

29 Charpignon ML, Vakulenko–Lagun B, Zheng B, et al. Causal inference in medical records and complementary systems pharmacology for metformin drug repurposing towards dementia[J]. Nat Commun, 2022, 13(1): 7652. DOI: 10.1038/s41467–022–35157–w.

30 Fang J, Zhang P, Zhou Y, et al. Endophenotype–based in silico network medicine discovery combined with insurance record data mining identifies sildenafil as a candidate drug for Alzheimer's disease[J]. Nat Aging, 2021, 1(12): 1175–1188. DOI: 10.1038/s43587–021–00138–z.

31 Yan C, Grabowska ME, Dickson AL, et al. Leveraging generative AI to prioritize drug repurposing candidates for Alzheimer's disease with real–world clinical validation[J]. NPJ Digit Med, 2024, 7(1): 46. DOI: 10.1038/s41746–024–01038–3.

32 Wu Y, Warner JL, Wang L, et al. Discovery of noncancer drug effects on survival in electronic health records of patients with cancer: a new paradigm for drug repurposing[J]. JCO Clin Cancer Inform, 2019, 3: 1–9. DOI: 10.1200/cci.19.00001.

33 Dickerman BA, García–Albéniz X, Logan RW, et al. Emulating a target trial in case–control designs: an application to statins and colorectal cancer[J]. Int J Epidemiol, 2020, 49(5): 1637–1646. DOI: 10.1093/ije/dyaa144.

34 Hubbard RA, Gatsonis CA, Hogan JW, et al. "Target trial emulation" for observational studies – potential and pitfalls[J]. N Engl J Med, 2024, 391(21): 1975–1977. DOI: 10.1056/NEJMp2407586.

35 Xie J, Wu EQ, Wang S, et al. Real–world data for healthcare research in China: call for actions[J]. Value Health Reg Issues, 2022, 27: 72–81. DOI: 10.1016/j.vhri.2021.05.002.

36 Zhang YF, Zhang H, Lipton Z, et al. Can transformers be strong treatment effect estimators?[EB/OL]. (2022–10–17) [2025–07–23]. https://doi.org/10.48550/arXiv.2202.01336.

37 Liu R, Chen PY, Zhang P. CURE: a deep learning framework pre–trained on large–scale patient data for treatment effect estimation[J]. Patterns, 2024, 5(6): 100973. DOI: 10.1016/j.patter.2024.100973.

收稿日期：2025 年 01 月 21 日 修回日期：2025 年 07 月 24 日
本文编辑：冼静怡 杨 燕