

# 《药物流行病学研究方法学指南（第2版）》系列 解读（16）：人工智能的应用探讨与辅助决策



吴昀效<sup>1, 2</sup>, 聂晓璐<sup>3</sup>, 颜济南<sup>1, 2</sup>, 詹思延<sup>1, 2, 4, 5</sup>, 孙 凤<sup>1, 2, 4, 6, 7, 8</sup>

1. 重大疾病流行病学教育部重点实验室（北京大学）（北京 100191）
2. 北京大学公共卫生学院流行病与卫生统计学系（北京 100191）
3. 国家儿童医学中心，首都医科大学附属北京儿童医院临床流行病与循证医学中心（北京 100045）
4. 北京大学医学部药品上市后安全性研究中心（北京 100191）
5. 北京大学第三医院临床流行病学研究中心（北京 100191）
6. 北京大学第三医院眼科（北京 100191）
7. 新疆医科大学中医学院（乌鲁木齐 830017）
8. 新疆石河子大学公共卫生学院（新疆石河子 832000）

**【摘要】**随着真实世界数据的日益积累与前沿算力支撑技术的显著提升，人工智能（AI）正深刻变革药物流行病学的研究范式。本文基于《药物流行病学研究方法学指南（第2版）》，系统探讨AI在药物流行病学研究中特定数据智能提取与深层数据洞察生成这两大核心应用方向。在数据提取方面，详细解析自然语言处理（NLP）与机器学习（ML）技术在非结构化医疗文本向结构化数据转化中的关键作用，同时强调，在具体临床应用场景中需审慎权衡精确度与召回率，并规范报告模型性能指标。在数据洞察层面，深入剖析ML应用于药物流行病学研究中复杂混杂偏倚控制、高维度临床预测模型构建及前沿概率表型分析中的方法学优势与现存局限。此外，本文聚焦方法学核心，重点梳理ML识别药物治疗效应异质性（HTE）的“基于风险”与“基于效应”双重建模路径。针对深度学习的“黑箱”特性，进一步探讨可解释人工智能（XAI）在辅助审查医疗风险、规避伦理问题中的重要性，及其揭示模型底层决策机制方面的局限性。本文旨在为研究者规范应用AI技术、提升药物流行病学研究质量及决策可靠性提供严谨的方法学支撑与实践参考。

**【关键词】**药物流行病学；方法学；指南；人工智能

**【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 (16): exploration of artificial intelligence applications and assisted decision making*

WU Yunxiao<sup>1, 2</sup>, NIE Xiaolu<sup>3</sup>, YAN Jinan<sup>1, 2</sup>, ZHAN Siyan<sup>1, 2, 4, 5</sup>, SUN Feng<sup>1, 2, 4, 6, 7, 8</sup>

1. Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China

DOI: 10.12173/j.issn.1005-0698.202603076

基金项目：国家自然科学基金面上项目（72474008）；国家自然科学基金国际（地区）合作与交流项目（72361127500）；海南省科学技术厅重点研发专项（ZDYF2024LCLH002）；国家药监局药品评价中心开放课题（CDR2024R01001）

通信作者：孙凤，博士，研究员，博士研究生导师，Email: sunfeng@bjmu.edu.cn

詹思延，博士，教授，博士研究生导师，Email: siyan-zhan@bjmu.edu.cn

2. Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China
  3. Center for Clinical Epidemiology & Evidence-based Medicine, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing 100045, China
  4. Center of Postmarketing Safety Evaluation, Peking University of Health Science Center, Beijing 100191, China
  5. Clinical Epidemiology Research Center, Peking University Third Hospital, Beijing 100191, China
  6. Department of Ophthalmology, Peking University Third Hospital, Beijing 100191, China
  7. School of Traditional Chinese Medicine, Xinjiang Medical University, Urumqi 830017, China
  8. School of Public Health, Shihezi University, Shihezi 832000, Xinjiang Uygur Autonomous Region, China
- Corresponding authors: SUN Feng, Email: sunfeng@bjmu.edu.cn; ZHAN Siyan, Email: siyan-zhan@bjmu.edu.cn

**【Abstract】** With the growing accumulation of real-world data and substantial advances in cutting-edge computing technologies, artificial intelligence (AI) is profoundly transforming the research paradigm of pharmacoepidemiology. Based on the *Guide on Methodological Standards in Pharmacoepidemiology (2nd edition)*, this paper systematically examines two core application domains of AI in pharmacoepidemiological research: targeted intelligent data extraction and the generation of in-depth data insights. In the domain of data extraction, this article elaborates on the pivotal role of natural language processing (NLP) and machine learning (ML) techniques in converting unstructured medical text into structured data, while emphasizing the need to carefully balance precision and recall in specific clinical application scenarios and to report model performance metrics in a standardized manner. In the domain of data insights, this article conducts an in-depth analysis of the methodological advantages and existing limitations of ML as applied to the control of complex confounding bias, the construction of high-dimensional clinical prediction models, and cutting-edge probabilistic phenotyping in pharmacoepidemiological research. Furthermore, with a focus on methodological fundamentals, this paper systematically reviews the dual modeling pathways of ML for identifying heterogeneity of treatment effect (HTE)—namely the "risk-based" and "effect-based" approaches. To address the "black-box" nature of deep learning, the importance of explainable AI (XAI) is further discussed in facilitating the review of medical risks and the avoidance of ethical concerns, as well as its inherent limitations in elucidating the underlying decision-making mechanisms of models. This paper aims to provide researchers with rigorous methodological support and practical guidance for the standardized application of AI technologies, with a view to enhancing the quality of pharmacoepidemiological research and the reliability of decision-making.

**【Keywords】** Pharmacoepidemiology; Methodology; Guidelines; Artificial intelligence

药物流行病学作为一门聚焦药品上市后安全监测、疗效评价与合理用药的交叉学科，其研究核心依赖多源医疗数据的挖掘与分析，而传统研究方法在处理非结构化数据、挖掘复杂变量关联时的局限性日益凸显<sup>[1-2]</sup>。随着数智化医疗的快速发展，电子健康档案（electronic health records, EHR）、医疗保险记录、疾病/医疗产品注册登记、可穿戴设备数据等多源异构数据呈现爆发式增长，为药物流行病学研究提供了丰富的数据支撑，但也对数据处理与分析技术提出了更高要求。人工智能（artificial intelligence, AI）技术，尤其是机器学习（machine learning, ML）与自然语言

处理（natural language processing, NLP），不仅极大地拓展了药物流行病学研究数据的可及性与分析维度，更为解决传统方法学中的高维混杂控制、异质性治疗效应（heterogeneity of treatment effect, HTE）精确估计等核心难题提供了创新解决路径<sup>[3-4]</sup>。在这一背景下，《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/修订工作于2024年启动<sup>[5]</sup>，2025年10月作为中国药学会团体标准（编号：T/CPHARMA 002-2025）正式发布实施。该指南前瞻性地纳入了AI在药物流行病学研究中的应用规范，标志着我国药物流行病学研究正式迈入

数智融合的新纪元。本文以指南第 2 版内容为核心，系统梳理 AI 在药物流行病学中的理论框架、应用场景及方法学挑战，并结合当前研究实践补充完善，为研究者与临床决策者提供指引。

## 1 AI驱动的非结构化数据提取与结构化转化

药物流行病学研究结果的可靠性高度依赖纵向连续临床数据的质量。然而，真实世界医疗场景中，患者主诉、影像与病理报告等海量关键临床信息，多以非结构化自由文本形式存在；传统人工查阅和编码的信息提取模式不仅耗时费力、成本高昂，还难以避免审核者主观差异导致的信息偏倚<sup>[6-7]</sup>。AI 技术，特别是 NLP 领域的技术突破，为非结构化文本向结构化、标准化且可直接用于统计分析的临床数据转化提供了高效的自动化路径。

### 1.1 非结构化医疗文本的处理技术与药物流行病学应用拓展

在药物流行病学研究中，基于 AI 的自动化数据提取核心任务是从内容复杂、篇幅冗长且充斥着大量医学缩写与非标准化表达的临床自由文本中，精准识别并提取研究预设的目标变量，而这一任务的实现主要依赖 NLP 的两项核心底层技术，即命名实体识别（named entity recognition, NER）与关系抽取（relation extraction, RE）<sup>[8]</sup>。

NER 的核心功能是从自由文本中定位、识别并分类预定义范畴的医学实体，例如药物通用名与商品名、特定疾病的诊断分类、实验室检验指标的异常状态、患者的症状与体征描述等。例如，在药物警戒领域，研究者利用 NER 从临床文本中提取药物名称及相关属性，如剂量、给药途径、强度和药品不良事件（adverse drug event, ADE）等<sup>[9-10]</sup>；Alfattni 等<sup>[9]</sup>评估了 NLP 和深度学习方法在从临床自由文本中提取和关联药物名称及相关属性的可行性；同样，medExtractR 系统也利用词典和正则表达式从 EHR 中提取药物信息，专注于特定药物以创建药物特异性研究数据集<sup>[10]</sup>。这些技术能够处理临床文本中常见的缩写、上下文相关的歧义以及时间序列的复杂性。然而，对于药物流行病学研究而言，仅提取孤立的医学实体无法支撑复杂的因果推断与关联分析，远不能满足研究

的核心需求。RE 技术则在实体识别的基础上，进一步挖掘并锚定不同实体间的语义关联、时序逻辑与因果指向，例如将特定时间窗内的药物暴露，与后续发生的 ADE、新发疾病适应证进行时序与因果层面的强关联绑定<sup>[11]</sup>。例如，Wei 等<sup>[12]</sup>构建了一个包含 NER 和关系分类组件的系统，用于从 EHR 中提取药物和相关的 ADE；此外，一项结合 ML 和规则算法的研究<sup>[13]</sup>旨在检测和识别出院总结中记录的药品不良反应（adverse drug reaction, ADR）的相关实体，该研究通过对非结构化出院总结文本的分析，提高了药物警戒的效率和准确性，为药物流行病学提供了更可靠的数据来源。尽管现有相关研究多聚焦于 EHR 的信息提取，但通过针对性的迁移学习与适当的微调策略（fine-tuning），该方法体系可拓展至其他类型的非结构化数据集，包括自发呈报系统、药品说明书以及社交媒体等；其应用还可延伸至系统评价与 Meta 分析的文献自动筛选流程，从而提升二次研究的执行效率与数据覆盖的全面性<sup>[14-15]</sup>。

一套高性能的临床 NLP 处理框架，通常涵盖文本去噪与预处理、词元化、基于医学领域本体库 [如 SNOMED CT、《监管活动医学词典》（*Medical Dictionary for Regulatory Activities*, MedDRA）] 的标准化实体映射，以及基于 Transformer 架构预训练语言模型等深度学习算法的时序关系与关联推断等核心环节<sup>[16]</sup>。这种端到端的处理范式，可实现对影像记录、病理报告甚至中医症状体征等非结构化医疗文本的大规模、高吞吐量自动化信息提取，对于药物警戒和药物安全性监测至关重要。针对传统 ADE 报告系统报告不足的缺陷，利用 NLP 技术可以从多模态数据中提取相关信息，以实现药物安全性的更全面监测；而针对 ADE 报告中药物编码耗时且面临挑战（如拼写错误、缩写或模糊商品名）的问题，有研究<sup>[17]</sup>验证了世界卫生组织全球药品名称词典（WHODrug Koda）在药物警戒数据库 VigiBase 中自动化药物编码的有效性和准确性，能够很好地处理模糊名称并提供有用的编码建议，展示了 AI 在自动化药物编码过程中的应用潜力。此外，AI 驱动的 NLP 技术也正应用于药物发现与开发中，大语言模型（large language model, LLM）能够改变 NLP 的格局，通过在靶点识别、药物重定

位、临床试验设计等方面提供帮助，从而加速新药的筛选和优化过程<sup>[18]</sup>。

## 1.2 模型性能评估的方法学权衡

指南第2版明确界定，AI在药物流行病学数据提取中的核心价值在于提升数据转化的效率与标准化水平，但需警惕技术应用中的方法学风险，其中模型性能评估的科学权衡是关键环节。指南第2版强调从非结构化文本中提取的数据并非绝对准确，不同算法模型（包括同一算法的不同参数设置）在相同数据提取任务中的性能表现可能存在显著差异，这种差异直接影响后续药物流行病学分析结果的可靠性，因此必须建立规范化的性能评估与场景化权衡体系。

在AI数据提取模型的性能评估中，精确度（precision）与召回率（recall）是两大核心且相互制约的指标，二者的权衡需严格根据药物流行病学的具体研究目标与临床应用场景，这也是指南第2版着重强调的方法学原则。精确度衡量模型预测为阳性样本中真正为阳性的比例，其核心在于最小化假阳性（false positives），避免将非目标实体错误提取为研究变量，从而降低变量错分和测量误差对后续分析的影响；召回率则衡量所有真实阳性样本被模型成功捕获的比例，其核心在于最小化假阴性（false negatives），避免遗漏关键临床信息（如潜在的ADE关联、罕见病相关实体等），尤其在药物警戒等需全面捕捉安全信号的场景中具有重要意义<sup>[19]</sup>。值得注意的是，精确度与召回率通常呈负相关，提高精确度往往会降低召回率，反之亦然，因此无法实现单一指标

的最大化，需基于研究场景进行针对性权衡。

在指南第2版基础上，本文总结了药物流行病学研究中主要应用场景的权衡策略示例（表1）。在自发ADR报告中识别医学概念并映射至标准化医学术语时，高召回率有助于发现潜在的药物安全信号，为早期预警提供支撑，但同时可能引入较多假阳性结果，增加后续人工复核的成本；而高精度、低召回率的模型虽能降低假阳性干扰，却可能遗漏低发生率但高风险的ADR信号，导致安全监测的局限性。此外，在可计算表型分析中，若研究目标为识别某一罕见病的全部潜在病例，需优先保证高召回率以避免病例遗漏；若研究目标为构建精准的病例队列用于药物疗效分析，则需优先保证高精度，减少非病例个体的混入。

除精确度与召回率的权衡外，指南第2版还明确要求，使用AI进行数据提取时，需详细报告模型的综合性能指标及验证结果，以保障研究的可重复性与可靠性。其中，F1分数（F1-score）作为精确度与召回率的调和平均数，能够综合反映模型的整体性能，是核心报告指标之一；同时需补充交叉验证（如10折交叉验证）、外部验证或独立人工复核结果，其中独立复核需由至少2名经过培训的临床或流行病学专业人员完成，以量化模型提取结果与金标准的一致性（如Kappa值）<sup>[24]</sup>。此外，对于多中心研究，还需报告模型在不同中心数据集中的性能稳定性，避免模型仅在单一数据集上过拟合而缺乏泛化能力。需特别强调的是，模型性能评估并非单一维度的指标比较，而是需结合数据特征、研究目标与临床意义

表1 药物流行病学研究中AI数据提取算法的精确度与召回率权衡策略

Table 1. Precision and recall trade-off strategies for AI data extraction algorithms in pharmacoepidemiological studies

药物流行病学应用场景	有限优化的性能指标	核心逻辑与方法学考量	潜在风险与局限性
ADR或安全信号的早期检测 <sup>[18, 20]</sup>	高召回率（high recall）	安全信号检测的根本目的在于发现潜在的、罕见的隐患，必须最大限度地扩大捕获网，宁可将有可疑病例纳入审核，也绝不能漏掉任何可能危及生命的安全信号	会产生大量的假阳性结果，导致后续人工临床审核、病历复核与因果评估的工作量呈指数级增加，造成资源消耗
药物有效性研究的严格队列构建与表型识别 <sup>[21]</sup>	高精度（high precision）	确保最终纳入研究队列的患者绝对满足预设的疾病诊断标准或特定暴露特征，坚决避免因错分偏倚（misclassification bias）稀释或扭曲真实的药物治疗效应 <sup>[22]</sup>	过于严苛的规则可能遗漏部分非典型临床表达的真实病例（低召回率），导致研究样本量大幅缩减，甚至可能引入选择偏倚（selection bias）
综合性临床特征数据的大规模自动化提取 <sup>[23]</sup>	F1分数的最优化	在不需要极端偏向某一方的常规流行病学建库任务中，要求算法在实体捕获率与准确率之间实现最优的数学平衡，为后续的多维度统计回归分析提供尽可能无偏的基线数据	在面对极度不平衡的临床数据集（如某类罕见病在全人群电子病历中的极低发生率）时，维持高水平的F1分数在算法优化上面临极大挑战

进行综合判断，这也是指南第 2 版对 AI 数据提取应用的核心规范要求。

## 2 AI赋能的数据洞察与药物流行病学辅助决策

数据洞察是 AI 赋能药物流行病学研究的关键价值所在，主要通过 ML 模型挖掘数据背后的复杂关联、生成临床决策支撑信息，具体涵盖混杂偏倚控制、临床预测模型构建与概率表型分析三大应用场景，各场景均需遵循指南明确的方法学规范。

### 2.1 ML辅助的倾向评分估计与混杂调整

混杂偏倚是药物流行病学观察性研究中影响结果真实性的主要问题之一，常用控制策略包括分层、多变量回归，以及基于倾向评分（propensity score, PS）的匹配、调整与加权等。实践中多采用 Logistic 回归模型估计 PS，但其存在明显局限性，如需手动选择协变量、难以拟合协变量与暴露因素之间的非线性关系及变量间的交互作用，尤其在高维协变量（如 EHR 中的多维度临床指标）场景下，传统方法的偏倚控制效果有限。ML 模型为解决上述问题提供了创新路径，其理论优势在于可自动完成协变量选择、高效捕捉非线性效应与复杂交互作用，无需预先设定变量关联形式，更适用于真实世界高维数据的混杂偏倚控制<sup>[25]</sup>。

目前，用于混杂偏倚控制的 ML 模型主要包括基于树的算法 [如随机森林、梯度提升树、贝叶斯加性回归树（Bayesian additive regression trees, BART）]、神经网络等，其中随机森林与梯度提升树因其可解释性相对较强、对异常值不敏感，在药物流行病学研究中应用最为广泛<sup>[26]</sup>。其原理是通过构建多棵决策树或迭代优化的回归树，充分挖掘协变量与暴露、结局之间的复杂关联，进而更精准地估计 PS 或直接调整混杂效应。例如，有研究利用真实世界数据（real world data, RWD）评估利奈唑胺与血液 ADE 之间的关联，采用随机森林分类模型自动纳入 EHR 中的数十项患者及临床相关变量（如初始医嘱、人口统计学信息、实验室检查结果、合并用药史等），显著提升混杂偏倚的控制效果<sup>[27]</sup>。

然而，ML 在混杂偏倚控制中的应用仍存在局限性，目前大多数相关研究仍基于合成数据或半合成模拟数据（plasmode data）开展，在 RWD

中的应用有待进一步拓展，其原因在于 RWD 存在数据缺失、异质性强、变量分布不均衡等问题，可能导致 ML 模型过拟合，影响混杂调整的稳定性；此外，ML 模型的协变量选择过程具有“黑箱”特性，难以明确解释各协变量的混杂贡献，不符合药物流行病学研究的可解释性要求。因此，在使用 ML 进行混杂偏倚控制时，需先对 RWD 进行严格的预处理（如缺失值填补、异常值剔除、变量标准化）；同时采用交叉验证、外部验证等方法评估模型的泛化能力；此外，需结合传统回归模型的结果进行对比分析，以验证 ML 模型混杂调整效果的可靠性，避免过度依赖单一模型导致的方法学偏倚<sup>[28]</sup>。

### 2.2 数据驱动的临床预测模型

数据驱动的临床预测模型是 AI 生成数据洞察的重要场景，通过利用多源 RWD 构建模型，预测药物相关的临床结局（如 ADE 发生风险、药物疗效、患者依从性等），为药物临床应用与安全监测提供精准决策支撑。ML 驱动的临床预测模型已在药物流行病学领域广泛应用，尤其在药物安全信号检测、个体化治疗决策等方面取得了一定成效，但其应用需严格遵循规范化的研究与报告流程。

与传统统计预测模型（如 Logistic 回归、Cox 比例风险回归）相比，ML 预测模型的优势在于可处理高维、异构数据，无需预设结局与预测变量之间的关联形式，能够自动捕捉复杂的非线性关系与变量交互作用，预测精度更高。在药物安全监测领域，基于 ML 的预测模型可通过整合 EHR、医疗保险数据、自发 ADR 报告等多源数据，预测特定人群使用某药物后发生 ADE 的风险，例如有研究<sup>[29]</sup>利用梯度提升决策树、极端梯度提升等 9 种 ML 模型结合患者年龄、性别、合并疾病、合并用药、肝肾功能等指标，预测癌症患者使用免疫检查点抑制剂的免疫介导性肝毒性风险，为临床用药筛选提供依据；在药物疗效预测领域，ML 模型可通过分析患者基因特征、临床指标与药物治疗反应的关联，预测患者对特定药物的疗效，为个体化用药方案制定提供支撑<sup>[30-31]</sup>。

指南第 2 版对 ML 预测模型的报告规范提出了明确要求：一是需清晰定义研究问题，明确预测目标（如预测某肿瘤药物治疗后的疾病进展风险）、目标人群、预测变量与结局变量，避免研

究问题模糊导致的模型构建偏差；二是需完整呈现预测模型的构建流程，包括数据来源、预处理方法、模型选择依据、参数调优过程、训练集与测试集的划分等，确保模型的可重复性；三是需客观评估模型性能，核心报告指标包括区分度[如受试者工作特征曲线下面积（area under the receiver operating characteristic curve, AUC）或一致性统计量（concordance statistic, C-statistic）]、校准度（如 Hosmer-Lemeshow 检验、校准曲线）以及临床实用性[如决策曲线分析（decision curve analysis, DCA）]，同时需通过外部验证或前瞻性验证评估模型的泛化能力，避免仅基于训练集评估导致的性能高估<sup>[32]</sup>。此外，ML 预测模型的应用需结合临床实际，不能单纯追求预测精度，还需考虑模型的可解释性与临床可行性，避免因“黑箱”特性导致临床决策者难以理解与应用<sup>[33]</sup>。

当前 ML 预测模型应用仍面临一些挑战，如部分模型存在过拟合现象、对小众人群（如罕见病患者、老年共病患者）的预测精度较低、多中心数据整合困难等，这些问题均需在未来研究中结合指南要求逐步优化，推动模型从“数据驱动”向“临床实用”转型。

### 2.3 概率表型分析与传统可计算表型的深层辨析

表型定义是药物流行病学研究中识别研究对象、构建分析队列和界定临床结局的重要基础。传统可计算表型（computable phenotyping）通过预设的诊断代码、实验室指标阈值或临床标准，根据 RWD 中的患者信息将其分类为“病例”或“非病例”，主要用于特定疾病患者的识别与队列构建，具有操作简单、可重复性强的优势，但存在灵活性不足、难以处理边界病例的局限性<sup>[34]</sup>。概率表型分析作为传统可计算表型的补充与延伸，是 AI 在药物流行病学表型定义中的重要应用方向，它能够提升表型定义的精准度与灵活性，尤其适用于复杂疾病、罕见病或共病患者的表型识别<sup>[35]</sup>；其基本原理是基于已标记的样本（金标准病例与非病例）训练 ML 模型，通过挖掘样本的多维度临床特征（如症状、体征、实验室指标、用药史等）与表型状态的关联，开发个性化的病例定义，最终输出每例患者属于某一表型的连续概率值（0~1），而非二元分类结果。

概率表型与传统可计算表型的主要差异在

于，传统可计算表型基于预设规则，属于“规则驱动”的表型定义方式，适用于疾病诊断标准明确、临床特征清晰的场景；概率表型基于 ML 模型，属于“数据驱动”的表型定义方式，适用于疾病异质性强、诊断标准不明确、边界病例较多的场景<sup>[36]</sup>。相较之下，传统可计算表型更适用于疾病定义明确、诊断标准较为统一的研究场景；而概率表型分析更适用于临床表现复杂、诊断边界模糊或存在明显异质性的研究对象，能够在一定程度上保留表型识别中的不确定性信息，并为后续敏感性分析或阈值优化提供依据。传统可计算表型已被用于耐药性高血压等疾病的识别与验证<sup>[21]</sup>；相比之下，概率表型方法则见于老年性痴呆、轻度认知障碍、阿尔茨海默病和帕金森病等复杂表型识别研究<sup>[36]</sup>。

### 3 药物HTE的智能识别

药物流行病学研究目标之一是揭示药物暴露与临床结局之间的因果关系，为临床合理用药提供依据。传统因果推断方法在处理复杂数据、估计个体化治疗效应时存在局限性，而 ML 技术的发展为因果推断提供了新的工具，尤其在 HTE 的识别中展现出独特优势<sup>[37]</sup>。HTE 是指不同患者对同一种药物治疗的反应存在差异，其识别对于个体化医疗和精准治疗具有重要意义，可帮助医生根据患者的具体情况选择最合适的治疗方案，提高治疗效果和减少不良反应。指南第 2 版将 ML 识别 HTE 的路径分为“基于风险的建模”与“基于效应的建模”两大类，两类路径的底层逻辑、适用场景与方法学特点存在显著差异，需根据研究目标选择合适的路径（图 1）。

基于风险的建模路径围绕着“药物治疗效果与结局事件基线风险相关”这一假设，其逻辑是先利用 ML 构建患者结局事件的风险预测模型，再通过比较不同风险分层下患者的效应差异，识别出获益最大的群体。该路径的优势在于可解释性较强，对复杂因果估计目标的依赖度较低，适用于结局事件明确、风险因素清晰的研究场景。例如，Xu 等<sup>[38]</sup>通过在随机对照试验中构建内部风险模型，对比了非参数与半参数方法在 HTE 分析中的应用，强调了基线风险分层在识别 HTE 中的必要性。基于效应的建模路径则直接以个体治疗效应（individual treatment effect, ITE）为预

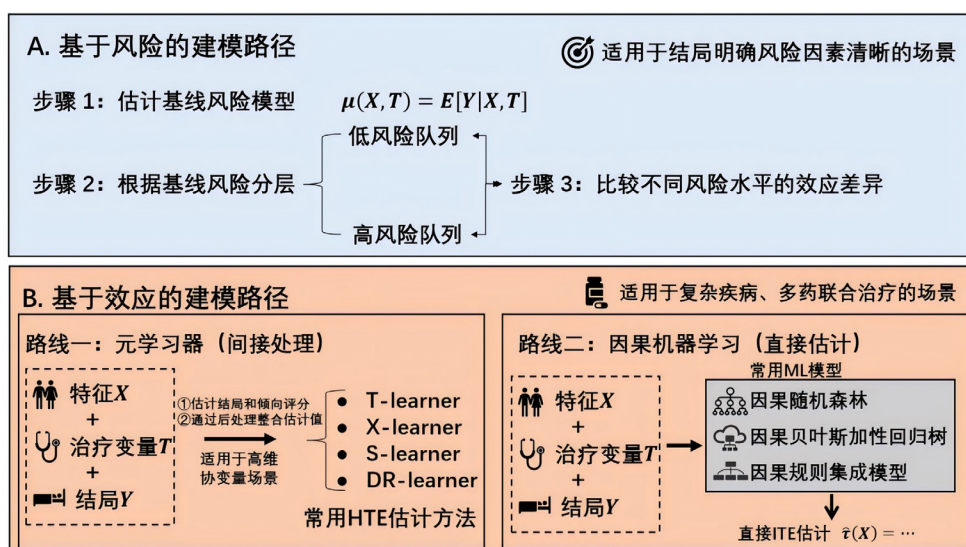


图1 药物HTE ML识别路径概念图

Figure 1. Conceptual diagram of ML pathways for identifying HTE

注:  $X$ : 特征/协变量, 代表患者个体的基线临床特征集合;  $T$ : 治疗/干预变量, 代表患者接受的干预状态;  $Y$ : 结局变量, 代表研究关注的临床结局指标;  $\mu$ : 结局期望, 代表模型对特定条件下的结局发生风险或数值的预测估计;  $\tau$ : 治疗效应, 代表干预措施带来的效应差值, 即个体化治疗效应或特定亚组的异质性治疗效应。

测目标, 其优势在于能更直接地捕捉交互作用, 适用于复杂疾病或多药联合治疗等场景。指南第 2 版将该路径细分为两条技术路线 (图 1)。第一条技术路线是基于元学习器 (Meta-learners) 的策略。该类方法将因果推断问题转化为一系类标准的监督学习问题。其中, S-Learner 将治疗变量作为普通特征处理; T-Learner 则分别对暴露组与非暴露组建模; X-Learner 在此基础上结合 PS 进行加权补偿, 能有效处理样本量极度不平衡或存在复杂交互作用的场景; DR-learner (双重稳健学习器) 则结合了 PS 与结局预测模型的优势, 具备“双重稳健性”, 即只要两个模型中有一个指定正确, 即可获得一致的效应估计。这类方法灵活性强, 广泛适用于高维协变量场景<sup>[37]</sup>。第二条技术路线是基于因果机器学习 (causal ML) 的直接估计策略。该路线利用特定算法直接最小化因果损失函数, 因果机器学习在此发挥了重要作用, 允许在个体层面估计治疗效应, 从而实现精准决策。例如, Galatro 等<sup>[39]</sup>探讨了基于树的算法在评估环境因素与健康结局因果关系中的应用, 包括因果随机森林 (causal random forest, CRF)、因果贝叶斯加法性回归树 (causal BART) 及因果规则集成模型, 旨在提升 HTE 估计性能的同时增强模型的可解释性。目前, 基于效应的路径仍存在局限, 特别是针对生存结局的因果估计方法虽已有随机生存森林等扩展,

但在处理竞争风险和临床应用方面尚需完善<sup>[37]</sup>。此外, ML 识别出的 HTE 结果必须结合临床实际, 通过外部验证与专家审核, 确保优势人群的界定具有临床统计学意义, 避免因算法偏差导致的决策失误。

#### 4 可解释人工智能与辅助决策伦理的深层挑战

ML 与深度学习的“黑箱”特性使其在药物流行病学应用中面临可解释性不足的困境, 不仅降低了临床决策的信任度, 更可能引发医疗风险与伦理争议<sup>[40]</sup>。为此, 指南第 2 版系统阐述了可解释人工智能 (explainable artificial intelligence, XAI) 的应用价值与局限性。当前 XAI 方法主要分为内在可解释模型与事后解释方法两类, 前者在模型构建时即融入透明度设计, 而后者则通过特定算法解析复杂模型 (如深度神经网络) 的决策过程, 是目前应用最广泛的技术路线。在代表性事后解释方法中, Shapley 加法解释 (Shapley additive explanations, SHAP) 值基于博弈论量化各特征对模型输出的全局或局部贡献度, 局部可解释模型无关解释 (local interpretable model-agnostic explanations, LIME) 方法则通过构建局部线性拟合以解释单一患者样本的预测逻辑, 高度适配个体化医疗场景。需要强调的是, XAI 通常解释的是预测关联或特征贡献, 而非药物与结

局之间的因果机制本身，因此其结果不能替代因果推断设计或临床判断。因此，XAI 仅能作为辅助工具，其解释结果必须结合临床经验进行综合校验，绝不能替代临床专家的独立判断<sup>[41]</sup>。

在推动 AI 技术落地的同时，药物流行病学研究必须直面并妥善解决 AI 辅助决策带来的伦理挑战，指南第 2 版将其归纳为算法公平性、数据隐私保护与责任界定。首先，模型极易继承训练数据中的历史偏差，导致对少数族裔、老年人或罕见病患者等特定群体的预测失准，加剧医疗资源分配不公；对此，研究者必须引入代表性强的多样化训练集，并动态评估与修正模型在不同特征子群中的性能。其次，鉴于研究依赖的 RWD 包含大量敏感信息，因此必须严格遵从数据保护法规，全面落实去标识化与匿名化等脱敏技术，以防数据泄露与滥用风险；另外，针对 AI 引发医疗不良事件时的责任边界模糊问题，应建立清晰的责任追溯机制，确立临床医生作为最终决策者的主体责任，同时界定模型开发者的合规义务。综上，AI 技术的临床应用必须坚守“以人为本”的原则，坚决防范因过度依赖算法而导致的“临床决策异化”，切实保障患者的健康与安全。

## 5 结语与展望：数智时代药物流行病学的守正与创新

本文基于指南第 2 版系统梳理了 AI 在药物流行病学中“数据提取”与“数据洞察”两大核心应用方向，深度剖析了其在非结构化 RWD 转化、混杂偏倚控制、概率表型分析、HTE 识别及可解释性 AI 辅助决策等场景中的方法学优势与现存局限。当前，AI 技术正推动该领域由“传统统计驱动”向“数智融合驱动”转型，极大突破了多维异构 RWD 的处理瓶颈。然而，作为发展阶段的新兴交叉领域，AI 应用尚未形成完善的方法学体系，仍面临严峻挑战，例如模型泛化能力受限，单中心数据衍生的算法在多中心、多样化 RWD 中的外部有效性亟待验证；深度学习的高性能与低可解释性矛盾尚未消弭；ML 在因果推断中的应用尚存技术壁垒，特别是针对生存结局的 HTE 识别及 AI 生成因果假设的验证流程仍不成熟；围绕数据隐私、算法公平性及责任界定的伦理合规体系仍亟待规范。

应对上述挑战，未来数智时代的药物流行病

学发展需聚焦于方法学标准化与深层次应用探索。首先，亟须依托指南建立涵盖模型构建、性能评估与结果报告的统一规范，以提升研究的透明度与可重复性；其次，需持续深化 XAI 技术与因果机器学习的交叉研发，在破解复杂模型“黑箱”属性的同时，完善 HTE 精准识别机制，为个体化诊疗提供更具公信力的因果证据支撑；另外，必须构筑严密的伦理审查与数据合规防火墙，保障技术的安全可持续发展。在此进程中，药物流行病学研究必须秉持“守正创新”的核心理念，既要坚守基于证据、追求真实可靠的流行病学方法学底线，同时也要深度融合 AI 优势以重塑数据处理与分析范式。唯有在严谨的方法学框架内规范应用数智技术，方能全面赋能药物上市后安全监测、临床合理用药与新药研发，最终切实推进“精准药学”的宏伟愿景。

**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 Sabaté M, Montané E. Pharmacoepidemiology: an overview[J]. *J Clin Med*, 2023, 12(22): 7033. DOI: [10.3390/jcm12227033](https://doi.org/10.3390/jcm12227033).
- 2 Qiao H, Chen Y, Qian C, et al. Clinical data mining: challenges, opportunities, and recommendations for translational applications[J]. *J Transl Med*, 2024, 22(1): 185. DOI: [10.1186/s12967-024-05005-0](https://doi.org/10.1186/s12967-024-05005-0).
- 3 Létinier L, Jouganous J, Benkebil M, et al. Artificial intelligence for unstructured healthcare data: application to coding of patient reporting of adverse drug reactions[J]. *Clin Pharmacol Ther*, 2021, 110(2): 392–400. DOI: [10.1002/cpt.2266](https://doi.org/10.1002/cpt.2266).
- 4 Capobianco E. High-dimensional role of AI and machine learning in cancer research[J]. *Br J Cancer*, 2022, 126(4): 523–532. DOI: [10.1038/s41416-021-01689-z](https://doi.org/10.1038/s41416-021-01689-z).
- 5 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): 121–135. [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](https://doi.org/10.12173/j.issn.1005-0698.202502028).
- 6 Zeng J, Gensheimer MF, Rubin DL, et al. Uncovering interpretable potential confounders in electronic medical records[J]. *Nature Commun*, 2022, 13(1): 1014. DOI: [10.1038/s41467-022-28546-8](https://doi.org/10.1038/s41467-022-28546-8).
- 7 Romanelli RJ, Schwartz NRM, Dixon WG, et al. The use of narrative electronic prescribing instructions in pharmacoepidemiology: a scoping review for the international society for pharmacoepidemiology[J]. *Pharmacoepidemiol Drug Saf*, 2021, 30(10): 1281–1292. DOI: [10.1002/pds.5211](https://doi.org/10.1002/pds.5211).

- 10.1002/pds.5331.
- 8 Fraile Navarro D, Ijaz K, Rezazadegan D, et al. Clinical named entity recognition and relation extraction using natural language processing of medical free text: a systematic review[J]. *Int J Med Inform*, 2023, 177: 105122. DOI: [10.1016/j.ijmedinf.2023.105122](https://doi.org/10.1016/j.ijmedinf.2023.105122).
  - 9 Alfattni G, Belousov M, Peek N, et al. Extracting drug names and associated attributes from discharge summaries: text mining study[J]. *JMIR Med Inform*, 2021, 9(5): e24678. DOI: [10.2196/24678](https://doi.org/10.2196/24678).
  - 10 Weeks HL, Beck C, McNeer E, et al. medExtractR: a targeted, customizable approach to medication extraction from electronic health records[J]. *J Am Med Inform Assoc*, 2020, 27(3): 407–418. DOI: [10.1093/jamia/ocz207](https://doi.org/10.1093/jamia/ocz207).
  - 11 Sharma T, Emmert–Streib F. Deep mining the textual gold in relation extraction[J]. *Artif Intell Rev*, 2024, 58(1): 34. DOI: [10.1007/s10462-024-11042-4](https://doi.org/10.1007/s10462-024-11042-4).
  - 12 Wei Q, Ji Z, Li Z, et al. A study of deep learning approaches for medication and adverse drug event extraction from clinical text[J]. *J Am Med Inform Assoc*, 2020, 27(1): 13–21. DOI: [10.1093/jamia/ocz063](https://doi.org/10.1093/jamia/ocz063).
  - 13 Tan HX, Teo CHD, Ang PS, et al. Combining machine learning with a rule–based algorithm to detect and identify related entities of documented adverse drug reactions on hospital discharge summaries[J]. *Drug Saf*, 2022, 45(8): 853–862. DOI: [10.1007/s40264-022-01196-x](https://doi.org/10.1007/s40264-022-01196-x).
  - 14 Delgado–Chaves FM, Jennings MJ, Atalaia A, et al. Transforming literature screening: the emerging role of large language models in systematic reviews[J]. *Proc Natl Acad Sci U S A*, 2025, 122(2): e2411962122. DOI: [10.1073/pnas.2411962122](https://doi.org/10.1073/pnas.2411962122).
  - 15 Sciurti A, Migliara G, Siena LM, et al. Compact large language models for title and abstract screening in systematic reviews: an assessment of feasibility, accuracy, and workload reduction[J]. *Res Synth Methods*, 2026, 17(2): 332–347. DOI: [10.1017/rsm.2025.10044](https://doi.org/10.1017/rsm.2025.10044).
  - 16 Zou H, Wang Y, Xiang K, et al. Knowledge–augmented large language model–based framework for adverse drug reaction analysis[J]. *Appl Soft Comput*, 2025, 185: 114025. <https://doi.org/10.1016/j.asoc.2025.114025>.
  - 17 Meldau EL, Bista S, Rofors E, et al. Automated drug coding using artificial intelligence: an evaluation of WHODrug Koda on adverse event reports[J]. *Drug Saf*, 2022, 45(5): 549–561. DOI: [10.1007/s40264-022-01162-7](https://doi.org/10.1007/s40264-022-01162-7).
  - 18 Liu Z, Roberts RA, Lal–Nag M, et al. AI–based language models powering drug discovery and development[J]. *Drug Discov Today*, 2021, 26(11): 2593–2607. DOI: [10.1016/j.drudis.2021.06.009](https://doi.org/10.1016/j.drudis.2021.06.009).
  - 19 Bomgni AB, Mbotchack Ngale CE, Aryal S, et al. NLPADADE: leveraging natural language processing for automated detection of adverse drug effects[A]. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine. IEEE, 2023: 4480–4487. DOI: [10.1109/bibm58861.2023.10385626](https://doi.org/10.1109/bibm58861.2023.10385626).
  - 20 Noguchi Y, Tachi T, Teramachi H. Detection algorithms and attentive points of safety signal using spontaneous reporting systems as a clinical data source[J]. *Briefings Bioinf*, 2021, 22(6): bbab347. DOI: [10.1093/bib/bbab347](https://doi.org/10.1093/bib/bbab347).
  - 21 McDonough CW, Babcock K, Chucru K, et al. Optimizing identification of resistant hypertension: computable phenotype development and validation[J]. *Pharmacoepidemiol Drug Saf*, 2020, 29(11): 1393–1401. DOI: [10.1002/pds.5095](https://doi.org/10.1002/pds.5095).
  - 22 Kan–Tor Y, Ness L, Szlak L, et al. Comparing the efficacy of anti–seizure medications using matched cohorts on a large insurance claims database[J]. *Epilepsy Res*, 2024, 201: 107313. DOI: [10.1016/j.eplepsyres.2024.107313](https://doi.org/10.1016/j.eplepsyres.2024.107313).
  - 23 McMaster C, Chan J, Liew DFL, et al. Developing a deep learning natural language processing algorithm for automated reporting of adverse drug reactions[J]. *J Biomed Inform*, 2023, 137: 104265. DOI: [10.1016/j.jbi.2022.104265](https://doi.org/10.1016/j.jbi.2022.104265).
  - 24 Khouri C, Revol B, Lepelley M, et al. A Meta–epidemiological study found lack of transparency and poor reporting of disproportionality analyses for signal detection in pharmacovigilance databases[J]. *J Clin Epidemiol*, 2021, 139: 191–198. DOI: [10.1016/j.jclinepi.2021.07.014](https://doi.org/10.1016/j.jclinepi.2021.07.014).
  - 25 Guo Y, Strauss VY, Català M, et al. Machine learning methods for propensity and disease risk score estimation in high–dimensional data: a plasmode simulation and real–world data cohort analysis[J]. *Front Pharmacol*, 2024, 15: 1395707. DOI: [10.3389/fphar.2024.1395707](https://doi.org/10.3389/fphar.2024.1395707).
  - 26 Cahn A, Raz I, Leiter LA, et al. Cardiovascular, renal, and metabolic outcomes of dapagliflozin versus placebo in a primary cardiovascular prevention cohort: analyses from DECLARE–TIMI 58[J]. *Diabetes Care*, 2021, 44(5): 1159–1167. DOI: [10.2337/dc20-2492](https://doi.org/10.2337/dc20-2492).
  - 27 Patel A, Doernberg SB, Zack T, et al. Predictive modeling of drug–related adverse events with real–world data: a case study of linezolid hematologic outcomes[J]. *Clin Pharmacol Ther*, 2024, 115(4): 847–859. DOI: [10.1002/cpt.3201](https://doi.org/10.1002/cpt.3201).
  - 28 Leite W, Zhang H, Collier Z, et al. Machine learning for propensity score estimation: a systematic review and reporting guidelines[J]. *Psychol Methods*, 2025. DOI: [10.1037/met0000789](https://doi.org/10.1037/met0000789).
  - 29 Wang B, Zhuang S, Lin S, et al. Analysis of risk factors for immune checkpoint inhibitor–associated liver injury: a retrospective analysis based on clinical study and real–world data[J]. *Hepatol Int*, 2025, 19(5): 1172–1186. DOI: [10.1007/s12072-025-10783-w](https://doi.org/10.1007/s12072-025-10783-w).
  - 30 Bang D, Koo B, Kim S. Transfer learning of condition–specific perturbation in gene interactions improves drug response prediction[J]. *Bioinformatics*, 2024, 40 (Suppl 1): i130–i139. DOI: [10.1093/bioinformatics/btae249](https://doi.org/10.1093/bioinformatics/btae249).
  - 31 Berchialla P, Lanera C, Sciannameo V, et al. Prediction of treatment outcome in clinical trials under a personalized medicine perspective[J]. *Sci Rep*, 2022, 12(1): 4115. DOI: [10.1038/s41598-022-07801-4](https://doi.org/10.1038/s41598-022-07801-4).
  - 32 Axford D, Sohel F, Abedi V, et al. Development and internal validation of machine learning–based models and external validation of existing risk scores for outcome prediction in patients

- with ischaemic stroke[J]. *Eur Heart J Digit Health*, 2024, 5(2): 109–122. DOI: [10.1093/ehjdh/zta073](https://doi.org/10.1093/ehjdh/zta073).
- 33 Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine[J]. *Nature Med*, 2022, 28(1): 31–38. DOI: [10.1038/s41591-021-01614-0](https://doi.org/10.1038/s41591-021-01614-0).
- 34 He T, Belouali A, Patricoski J, et al. Trends and opportunities in computable clinical phenotyping: a scoping review[J]. *J Biomed Inform*, 2023, 140: 104335. DOI: [10.1016/j.jbi.2023.104335](https://doi.org/10.1016/j.jbi.2023.104335).
- 35 Banda JM, Shah NH, Periyakoil VS. Characterizing subgroup performance of probabilistic phenotype algorithms within older adults: a case study for dementia, mild cognitive impairment, and Alzheimer's and Parkinson's diseases[J]. *JAMIA Open*, 2023, 6(2): ooad043. DOI: [10.1093/jamiaopen/ooad043](https://doi.org/10.1093/jamiaopen/ooad043).
- 36 Ding L, Mane R, Wu Z, et al. Data-driven clustering approach to identify novel phenotypes using multiple biomarkers in acute ischaemic stroke: a retrospective, multicentre cohort study[J]. *EClinicalMedicine*, 2022, 53: 101639. DOI: [10.1016/j.eclinm.2022.101639](https://doi.org/10.1016/j.eclinm.2022.101639).
- 37 Ling Y, Upadhyaya P, Chen L, et al. Emulate randomized clinical trials using heterogeneous treatment effect estimation for personalized treatments: methodology review and benchmark[J]. *J Biomed Inform*, 2023, 137: 104256. DOI: [10.1016/j.jbi.2022.104256](https://doi.org/10.1016/j.jbi.2022.104256).
- 38 Xu E, Vanghelof J, Wang Y, et al. Outcome risk model development for heterogeneity of treatment effect analyses: a comparison of non-parametric machine learning methods and semi-parametric statistical methods[J]. *BMC Med Res Methodol*, 2024, 24(1): 158. DOI: [10.1186/s12874-024-02265-8](https://doi.org/10.1186/s12874-024-02265-8).
- 39 Galatro D, Di Nardo A, Pai V, et al. Considerations for using tree-based machine learning to assess causation between demographic and environmental risk factors and health outcomes[J]. *Environ Sci Pollut Res Int*, 2024, 31(51): 60927–60935. DOI: [10.1007/s11356-024-35304-4](https://doi.org/10.1007/s11356-024-35304-4).
- 40 Gupta J, Seeja KR. A comparative study and systematic analysis of XAI models and their applications in healthcare[J]. *Arch Comput Methods Eng*, 2024, 31(7): 3977–4002. DOI: [10.1007/s11831-024-10103-9](https://doi.org/10.1007/s11831-024-10103-9).
- 41 Noor AA, Manzoor A, Mazhar Qureshi MD, et al. Unveiling explainable AI in healthcare: current trends, challenges, and future directions[J]. *WIREs Data Min Knowl*, 2025, 15(2): e70018. DOI: [10.1002/widm.70018](https://doi.org/10.1002/widm.70018).

收稿日期: 2026年03月19日 修回日期: 2026年04月05日  
本文编辑: 洗静怡 杨燕