·方法学指南解读·

《药物流行病学研究方法学指南（第2版）》系列解读（18）：治疗效应异质性的分析策略与临床应用

黄 涛[1]，沈卓恒[1]，胥 洋[1,2]

1. 北京大学药学院药事管理与临床药学系（北京 100191）
2. 重大疾病流行病学教育部重点实验室（北京大学）（北京 100191）

【摘要】治疗效应异质性（HTE）是指治疗效应在不同患者特征、疾病状态或临床背景下存在非随机差异。随着精准医学、真实世界证据和药品监管科学的发展，HTE分析已成为连接群体平均证据与个体化临床决策的重要方法学工具。本文基于《药物流行病学研究方法学指南（第2版）》，对HTE的基本概念、效应估计量、效应测量指标和因果识别假设进行解读；重点介绍亚组分析、风险建模和效应建模三类分析策略的基本原理、适用场景、实施步骤和典型案例；并讨论HTE研究在真实世界数据中的实施流程、可信度评价、规范报告与临床转化。

【关键词】治疗效应异质性；药物流行病学；真实世界数据；亚组分析；风险建模；效应建模；因果机器学习；精准医学

【中图分类号】R181.3+5 **【文献标志码】**A

Guide on Methodological Standards in Pharmacoepidemiology (2nd edition) and their series interpretation (18): analysis strategies and clinical applications of heterogeneity of treatment effect

HUANG Tao[1], SHEN Zhuoheng[1], XU Yang[1,2]

1. Department of Pharmacy Administration and Clinical Pharmacy, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China
2. Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China
Corresponding author: XU Yang, Email: xuyang_pucri@bjmu.edu.cn

【Abstract】Heterogeneity of treatment effect (HTE) refers to non-random variation in the magnitude or direction of treatment effects across patient characteristics, disease states, or clinical contexts. With the development of precision medicine, real-world evidence, and regulatory science, HTE analysis has become an important methodological tool for bridging population-average evidence and individualized clinical decision-making. Based on the framework of the *Guide on Methodological Standards in Pharmacoepidemiology (2nd edition)*, this article interprets the core concepts, estimands, scale dependence, and causal identification assumptions of HTE. It introduces three approaches of subgroup analysis, risk modelling, and effect modelling, with emphasis on their principles, applicable scenarios, implementation steps, and representative case

DOI: 10.12173/j.issn.1005-0698.202606071
基金项目：国家自然科学基金青年科学基金项目（82304245）
通信作者：胥洋，副研究员，博士研究生导师，Email：xuyang_pucri@bjmu.edu.cn

studies. Furthermore, this article discusses implementation workflows, credibility assessment, reporting, and clinical translation in real-world data studies.

【Keywords】Heterogeneity of treatment effect; Pharmacoepidemiology; Real-world data; Subgroup analysis; Risk modelling; Effect modelling; Causal machine learning; Precision medicine

随机对照试验（randomized controlled trial，RCT）通常以平均治疗效应（average treatment effect，ATE）作为主要结果。平均效应对于判断药物在目标总体中的总体获益或风险具有重要意义，但并不能充分回答临床实践中“哪类患者可能获益有限或面临风险”或“特定治疗用于具体病人时可能的效果是什么”的问题[1-2]。药物流行病学研究同样面临这一挑战。尽管真实世界数据（real-world data，RWD）为评价药物效应在不同人群中的分布提供了重要机会，且往往能够覆盖如高龄、多病共存、肾功能不全、复杂用药或医疗可及性差异显著等各类临床试验代表性不足的患者，但若只报告平均效应，同样会掩盖重要异质性信息[3]。

更为重要的是，RWD中的治疗效应异质性（heterogeneity of treatment effect，HTE）研究并不天然可靠，患者所接受的治疗分配常与疾病严重程度、既往治疗、医生处方偏好、医保政策和患者健康行为相关[4-5]。当总体因果效应估计受到适应证混杂、选择偏倚、永恒时间偏倚、时间依赖混杂或测量误差影响时，进一步开展HTE研究可能只是在有偏估计基础上产生更复杂的错误结论。因此，HTE分析首先是一个流行病学和因果推断问题，其后才是统计建模问题[6-7]。本文重点阐释了HTE方法的基本原理、适用场景、实施步骤和真实世界研究中的应用。

1 HTE的基本概念

1.1 HTE的流行病学内涵

HTE是指治疗效应在不同患者特征或临床背景下存在非随机差异（图1）[8]，该概念与流行病学中的效应修饰（effect modification）密切相关[9]。若治疗效应在不同水平的某变量上呈现差异，则称该变量为效应修饰因素。通常，效应修饰因素应为治疗前可测量的患者特征，例如年龄、性别、遗传标志物、疾病严重程度、合并症、肾功能、既往用药或药物代谢相关基因型等。若某一变量本身是第二种干预或治疗后的结果，例如后续联合用药、治疗依从性或治疗后实验室指标，则不宜简单作为基线效应修饰因素处理，而应重新定义为联合暴露、动态治疗策略或中介问题。

如图1所示，相同药物治疗下，不同特征的患者治疗效应不同。在药物流行病学中，HTE研究的意义已超出单纯识别药物对“哪类患者有效”的范围，而是沿着一条递进的证据链展开：首先识别“哪类患者更可能获益”，进而判断“哪类患者可能面临风险”，在此基础上综合“哪类患者的获益-风险比最优”，最终指向“哪些患者应优先获得治疗”。这一从疗效到决策的递进逻辑，使HTE分析与药品监管、标签更新、风险管理计划、医保支付和临床指南制定产生了直接联系。

1.2 效应估计量

药物流行病学领域中的HTE分析通常基于潜在结果框架（potential outcomes framework）[10]。为便于后续讨论，表1统一给出本文使用的符号及其含义。在该符号体系下，HTE效应估计目标可从群体到个体形成三个层级，如图2所示。

第一层级是ATE，其回答的问题是“治疗在目标人群总体中的平均疗效”，见图2，表达式为：

$$ATE = E[Y^1 - Y^0]$$

第二层级是条件平均治疗效应（conditional average treatment effect，CATE），其回答的问题是“治疗具有特定特征组合$X=x$的患者群体的疗效”，见图2，表达式为：

$$CATE(x) = \mathrm{E}[Y^1 - Y^0 | X = x]$$

第三层级是理论上的个体治疗效应（individualized treatment effect，ITE），其回答的问题是“治疗在单个患者中的疗效”，见图2。然而，由于同一患者无法同时被观察到接受治疗与不接受治疗两种状态，因此真实的ITE是无法直接观测的，表达式为：

$$ITE = Y_i^1 - Y_i^0$$

1.3 效应测量指标

理解和比较疗效，首先要明确在何种测量指

标上进行。同一治疗在不同效应测量指标上可能会给出截然相反的HTE结论，这是HTE分析中最容易被忽视却又是最根本的问题[11]。

以二分类结局为例，设$p(a,x)=E(Y^a|X=x)$是特征组合为$X=x$的患者在治疗a下的结局风险，那么绝对效应测量指标，如风险差（risk difference，RD）和相对效应测量指标，如风险比（risk ratio，RR）的表达式如下：

$$RD(x)=p(1,x)-p(0,x)$$

$$RR(x)=p(1,x)/p(0,x)$$

若某药物在所有患者中使风险比降低20%，那么基线风险为5%的患者绝对风险降低约1%，而基线风险为30%的患者绝对风险降低约6%；虽然以相对效应测量指标RR衡量治疗效应相同，但以绝对测量指标RD衡量药物治疗效应却呈现出重要的异质性。因此在HTE研究中，通常推荐

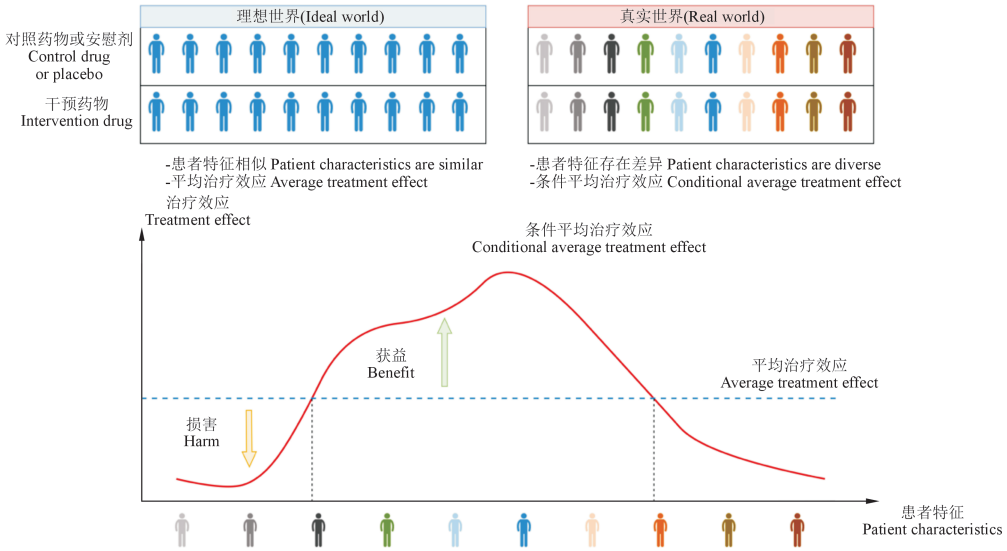

图1 治疗效应异质性示意图

Figure 1. Schematic diagram of heterogeneity of treatment effect

表1 潜在结果框架符号定义

Table 1. Symbol definition of potential outcomes framework

符号	含义
i	个体编号，$i=1,2,\cdots,n$
A	治疗策略分配指标：$A=1$接受研究治疗，$A=0$接受对照治疗
Y^1	同一患者若接受研究治疗时的潜在结局
Y^0	同一患者若接受对照治疗时的潜在结局
X	治疗前基线特征向量（可包含人口学、临床、治疗相关变量）
$E[\cdot]$	在目标总体上的期望

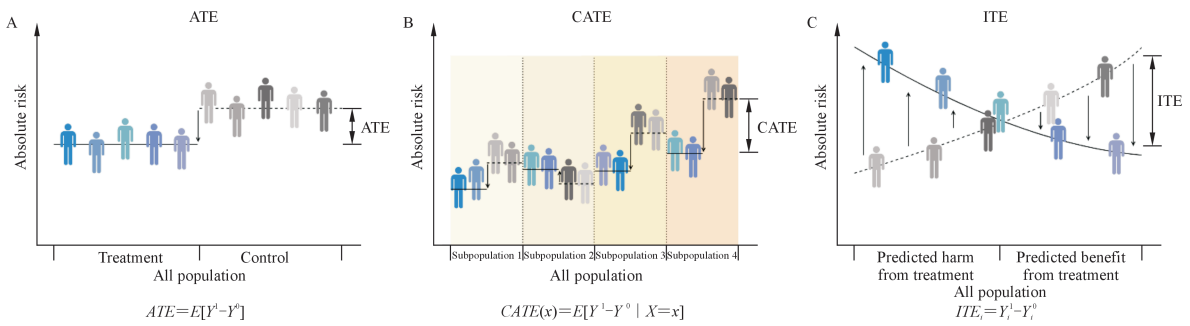

图2 不同HTE建模策略下的效应估计量

Figure 2. Estimands of effect under different HTE modeling strategies

注：ATE. 平均治疗效应；CATE. 条件平均治疗效应；ITE. 个体治疗效应。

以RD、限制平均生存时间差等绝对效应测量指标作为主要效应测量指标，同时以RR、危险比（hazard ratio，HR）或优势比（odds ratio，OR）等相对效应测量指标作为补充。

同时，对于二分类效应修饰因素Z，绝对效应测量指标上的效应修饰可表现为$RD(Z=1)-RD(Z=0)\neq 0$，相对效应测量指标上的效应修饰可表现为$RD(Z=1)/RD(Z=0)\neq 1$。因此HTE研究不应仅报告交互作用的P值，还应报告不同亚组或风险层内的结局风险、绝对效应和相对效应。

1.4 因果识别假设

HTE分析是因果推断问题，而不是普通预测问题。普通预测模型回答“在给定治疗和患者特征时，结局风险是多少”；HTE分析回答“同一类患者在不同治疗策略下的潜在结局有何差异”。因此，HTE分析必须满足一致性（consistency）、可交换性（exchangeability）和正性（positivity）等基本假设。

一致性即观察到的事实结局与潜在结局之间定义一致：当个体实际接受的治疗策略A为a时，观察到的Y应等于该个体的潜在结局Y^a。要求：①治疗策略定义足够明确，治疗起始、剂量、给药途径、持续时间、联合用药和宽限期等要素均应被规定；②同一治疗定义下不存在无法区分的多种实现版本，否则Y^a将不再指向一个确定的潜在结局。

可交换性要求在给定治疗前协变量X后，治疗策略分配与潜在结局独立：

$$(Y^1, Y^0)\perp A|X$$

在RCT中，该假设主要由随机化保证。在RWD中，则需要通过目标试验仿真、活性对照新用药者（active comparator new user，ACNU）设计、倾向评分、标准化、加权或双稳健估计等方法尽量实现可比性[12-14]。

正性要求每一类患者均有接受不同治疗策略的可能性：

$$0<P(A=1|X=x)<1$$

正性假设对HTE研究尤其关键。高维建模可以为许多特征组合生成治疗效应预测，但若某些特征组合下几乎没有治疗组或对照组样本，其估计主要依赖模型外推，因果解释有限。RWD中的HTE研究必须报告重叠性、协变量平衡、极端权重和有效样本量。

上述三条假设存在逻辑次序：一致性是反事实定义的基石，无一致性则所有估计均失去意义；可交换性赋予反事实比较以因果含义；正性则决定该比较在给定数据中是否可被可靠执行。

2 HTE分析策略

根据HTE的预测方法（the Predictive Approaches to Treatment Effect Heterogeneity，PATH）声明[8]和RWD中HTE研究[4-5]的最新方法学框架，HTE研究可概括为三种分析策略（图3）：亚组分析（subgroup analysis）、风险建模（risk modelling）和效应建模（effect modelling）。三者并非简单的技术替代关系，而是对应不同研究目标、不同数据条件和不同证据用途。

2.1 亚组分析

2.1.1 适用场景

亚组分析是最经典的HTE方法，适用于存在明确先验假设、机制依据或监管关注的情境。例如，基于年龄、性别、基因型、疾病严重程度或既往病史定义亚组，评价药物在不同人群中的疗效或安全性差异。药物研发中的生物标志物分层、上市后安全性信号验证、特殊人群用药评价和临床指南中的人群分层推荐，均以亚组分析为主要HTE分析工具。

亚组分析按目的不同可以分为确证性亚组分析和探索性亚组分析，尤其适合回答确证性问题[15]，即验证某个预设变量是否修饰治疗效应。事后探索性分析发现的亚组仅应作为假设生成，而非确证性结论[16-17]。若研究者在大量变量中事后寻找“显著亚组”，则假阳性风险和选择性报告风险明显升高。

2.1.2 实施流程

亚组分析的实施可分为四步。第一步，基于临床机制、既往证据或监管需求预先指定少数效应修饰变量，将协变量X拆分为关注的效应修饰因子Z和混杂/预后因素L。第二步，在每个亚组内估计治疗效应，并同时报告绝对效应和相对效应。第三步，在模型中纳入治疗、亚组变量及其交互项：

$$g\{E(Y|A,Z,L)\}=\beta_0+\beta_A A+\beta_Z Z+\beta_{AZ}(A\times Z)+\gamma L$$

其中$g(\cdot)$为连接函数，β_{AZ}表示特定模型和效应量上的交互作用。第四步，对交互结果进行可信度评价[18]，包括是否预设、是否存在多重比

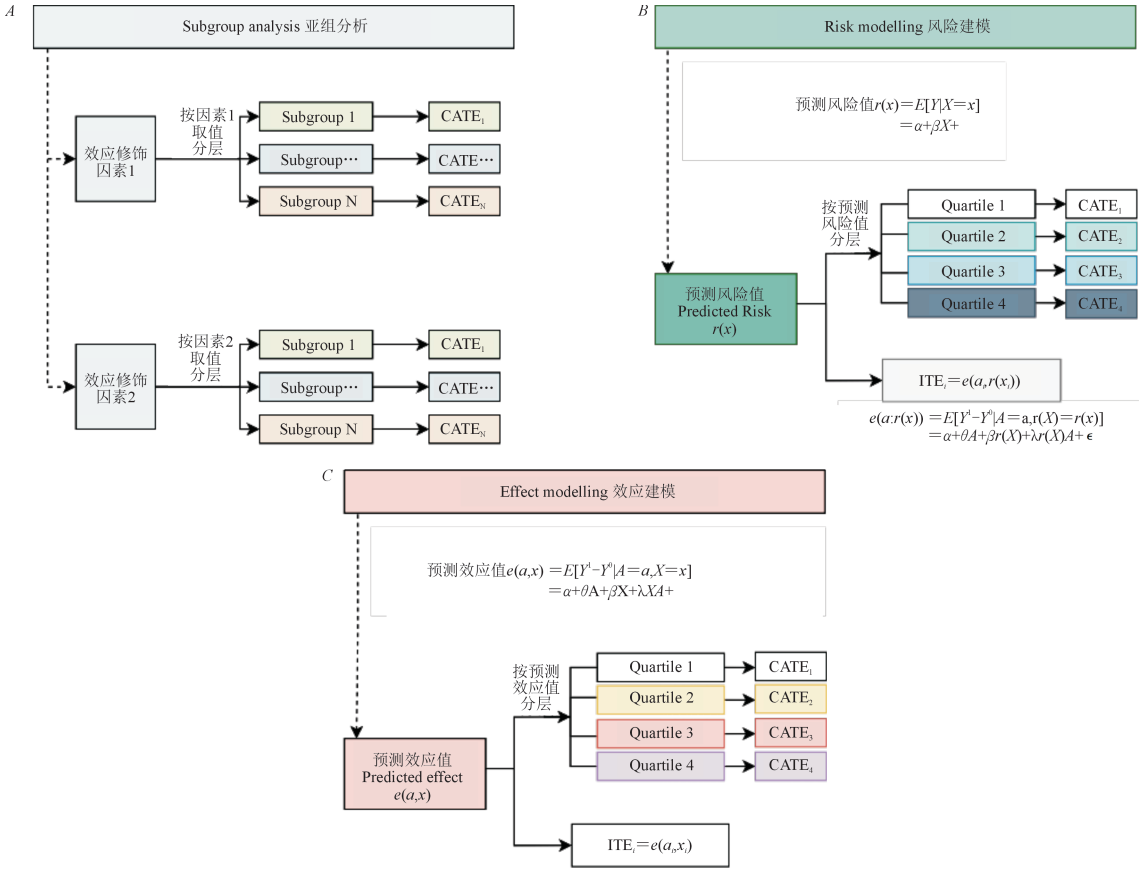

图3 HTE分析的三种分析策略

Figure 3. Three analysis strategies for HTE analysis

注：A. 亚组分析；B. 风险建模；C. 效应建模。CATE. 条件平均治疗效应；ITE. 个体治疗效应；Quartile. 四分位数。

较、效应方向是否与先验证据一致、效应大小是否具有临床意义、是否可被外部数据复现等。

2.1.3 案例

肿瘤靶向治疗是亚组分析推动精准医学的典型场景。在未经分子标志物选择的非小细胞肺癌患者中，某些靶向药物的总体疗效可能并不突出，但在表皮生长因子受体激活突变阳性患者中疗效显著[19-21]。案例说明，ATE可能掩盖具有明确生物学机制的获益人群。药物安全性研究中，若不良反应集中于特定药物代谢基因型或肾功能不全患者，亚组分析可为药品标签更新和风险管理提供依据。

2.1.4 优势与局限

亚组分析的优势是透明、易解释、便于临床和监管沟通。其局限包括：①统计效能不足：交互作用的检验效能远低于主效应，中等规模的亚组效应差异往往无法检出；②多重比较：亚组越多，假阳性风险越高；③亚组定义任意：同一连续变量取不同切点，结论可能翻转；④维度受限：每次只能处理少数变量，无法应对高维效应修饰情景；⑤同一患者可能同时属于多个亚组（如65岁以上女性和肾功能不全），当不同亚组结论不一致时，难以判断哪个结果最适用于该患者[16-18]。

2.2 风险建模

2.2.1 适用场景

风险建模关注患者基线结局风险与绝对治疗获益之间的关系，其方法论基础在于认为治疗的相对效应在所有患者中大致恒定，基线风险的差异会导致绝对获益存在异质性。因此，风险建模的核心问题不是“哪些变量直接修饰药物反应机制”，而是“哪些患者因基线风险高而获得更大的绝对获益”。

风险建模适用于以下场景：①总体治疗有效，即ATE明确指向获益；②患者间结局风险存在显著差异；③存在经良好验证的风险预测模型；④假定相对效应近似同质，即RR在患者间变化有限。在这一假设下，设$r_0(x)=\mathrm{E}(Y^0|X=x)$为

接受对照治疗策略特征为$X=x$的患者的基线风险，RR为近似恒定的风险比，则不同基线风险层内的绝对RD可近似表达为：

$$RD(x)\approx r_0(x)\cdot(RR-1)$$

当$RR<1$时，基线风险越高，绝对获益越大。因此，风险建模适合用于确定治疗优先级、医保支付人群、指南推荐顺序和获益-风险评估等决策场景。

2.2.2 实施流程

风险建模一般包括三步。第一，建立或选择风险预测模型。该模型应使用治疗前变量，如年龄、合并症、疾病严重程度、实验室指标、既往用药和医疗利用情况。若已有经良好验证的风险预测模型，应优先考虑已验证的风险预测模型；若需在研究数据中开发内部模型，应评估模型的区分度和校准度。第二，根据预测风险值进行分层。可使用四分位数、五分位数、临床阈值或连续风险评分。分位数分层有助于保证样本量，临床阈值则更利于解释和转化。第三，在各风险层内估计治疗效应，并同时报告绝对效应和相对效应。若风险层越高，绝对获益越大，且差异跨越临床决策阈值，则可认为存在具有临床意义的风险驱动HTE。此时应注意，若相对效应在各风险层内明显不一致，则恒定RR的假设可能不成立，风险建模的简化逻辑需要重新审视。

2.2.3 案例

PATH声明以RITA-3随机试验作为风险建模示例。该研究比较了非ST段抬高急性冠脉综合征患者的早期介入策略与保守治疗策略，并使用由年龄、性别、糖尿病、既往心肌梗死、吸烟、心率、ST段压低、心绞痛严重程度和左束支传导阻滞等临床特征构建的风险模型进行基线风险分层。结果显示，高风险患者接受早期介入策略的绝对获益更为明显，而低风险患者获益有限，可能因程序相关风险抵消获益。该案例体现了风险建模的临床价值：即使总体ATE提示治疗有效，真正需要优先接受治疗的人群也可能集中在高风险层[22]。

在药物流行病学中，类似思路可用于识别最可能从心肾保护药物、抗凝治疗、降脂治疗或高成本创新药中获得净获益的人群[23-24]，也可用于安全性评价，例如识别基线出血风险较高而不宜接受某种治疗的人群。

2.2.4 优势与局限

风险建模的优势是临床意义明确、可解释性强、能够整合多个预后因素，并可直接转化为绝对获益、治疗需要数（number needed to treat，NNT）、伤害需要数（number needed to harm，NNH）和净获益。这些特点使其在监管和医保分层决策中特别受青睐。其局限是风险评分主要反映预后风险，而不一定反映治疗反应机制。这一局限的根本原因在于风险建模通过预后风险逼近绝对获益差异，本质上是将预后异质性作为效应异质性的代理变量，因此带来了两个问题：①若治疗效应主要由非预后因素驱动（如特定药物代谢基因型），可能遗漏关键异质性来源；②该方法适合识别“谁的绝对获益更大”，但不一定能回答“为什么这些患者反应不同”[25-27]。

2.3 效应建模

2.3.1 适用场景

效应建模直接估计治疗效应如何随多个患者特征变化，其目标是直接估计CATE。与亚组分析相比，效应建模能够同时处理多个变量；与风险建模相比，它不局限于基线结局风险，而是尝试识别更复杂的治疗反应差异[28-29]。效应建模适用于样本量较大、协变量丰富、预期存在非线性或高阶交互的研究场景，如大型RCT、电子健康记录、医保数据库、登记队列和多组学数据研究。效应建模方法众多[30-31]，可按其与方法结构的关系分为两类：模型特定方法（model-specific）和模型无关方法（model-agnostic）。

2.3.2 模型特定方法

模型特定方法是指CATE的表达和解释依赖特定模型结构或算法。在选择某一特定方法的同时，也选择了CATE函数的形式约束。典型方法包括回归交互模型、惩罚回归、因果树、因果森林、贝叶斯加性回归树（Bayesian additive regression trees，BART）和贝叶斯因果森林（Bayesian causal forest，BCF）。

①回归交互模型是最简单的模型特定方法，通过在参数回归中纳入治疗、候选效应修饰变量及其交互项实现[32]，适合少数预设效应修饰因素的验证性分析，但难以捕捉非线性。

②惩罚回归［如带交互项的Least Absolute Shrinkage and Selection Operator（LASSO），即最小绝对收缩和选择算子或弹性网络］适合候选交互

项较多但仍希望筛选出少数重要项的场景。其估计结果可解释性强，但对交互项的事前编码敏感。

③因果树通过递归分裂寻找治疗效应差异最大的患者子群，输出以决策树形式呈现，结果直观但稳定性有限，分裂点依赖样本扰动[33]。因果森林通过集成多棵因果树提高估计的稳定性，能够捕捉复杂非线性和高阶交互，并提供逐点置信区间。其“诚实估计”（honest estimation）策略，即使用不同样本进行分裂和估计，是减少过拟合的关键[34]。

④BART 和 BCF 使用贝叶斯树模型估计潜在结局和治疗效应[35-36]，其中 BCF 通过区分预后函数和治疗效应函数，有助于减少预后风险强信号对效应异质性估计的干扰。

模型特定方法的优势是能够刻画复杂效应修饰模式，部分方法具有较强解释性。其局限是结果依赖模型设定、变量编码、调参策略和样本结构，且可能出现过拟合和虚假 HTE。

2.3.3 模型无关方法

模型无关方法并不意味着不使用模型，而是指研究者首先定义目标因果估计量，再使用不同模型估计所需组成部分。这种策略使得估计目标与具体算法解耦，可以灵活替换不同的学习器。典型方法包括两类策略：①基于标准化/加权的 CATE 估计；②基于元学习器（Meta-learner）的 CATE 估计。

前者包括标准化、逆概率加权（inverse probability weighting，IPW）、增强逆概率加权（augment inverse probability weighting，AIPW）和目标最大似然估计（target maximum likelihood estimation，TMLE）。其思路与 ATE 估计一致，只是将估计目标从 $E[Y^1-Y^0]$ 推广为 $E[Y^1-Y^0|X=x]$。以基于标准化的 CATE 估计为例，若 $m_a(x)$ 为治疗 a 下的结局函数，则特征为 $X=x$ 的患者的 CATE 可通过 $m_1(x)-m_0(x)$ 估计。

Meta-learner 框架则将 CATE 估计拆解为若干基础预测任务，再组合得到效应估计[37]。其核心思想是将目标因果估计量与具体预测算法解耦：研究者先明确需要估计的 CATE，再根据不同框架分别构建结局预测模型、倾向评分模型、反事实结局、残差或伪结局，最终得到 ITE 估计。不同 Meta-learner 的差异主要体现在基础任务的拆解方式上：S-learner（Single-learner）使用单一结局模型同时纳入治疗状态和协变量；T-learner（Two-learner）分别构建治疗组和对照组结局模型；X-learner 通过交叉预测反事实结局并以差值训练 CATE 模型；R-learner 通过结局和治疗残差化估计 CATE；DR-learner（Doubly robust-learner）则构造双稳健伪结局后训练 CATE 模型[38]。五种常见元学习器的基本流程见图 4，其方法学特性见表 3。值得注意的是，因果森林、BART 和 BCF 等模型特定方法可以用于元学习器中的基础预测任务。

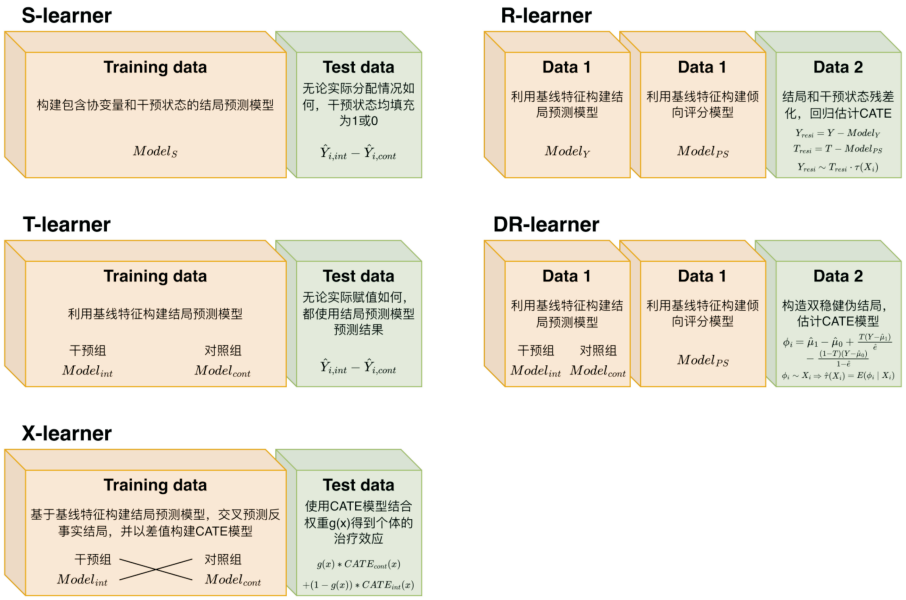

图 4 常见元学习器框架的基本流程图例

Figure 4. Basic flowchart of common Meta-learner frameworks

注：CATE. 条件平均治疗效应；int. 干预研究治疗；cont. 对照治疗；PS. 倾向性得分。

需注意，DR-learner和R-learner在理论上具有更优的收敛性质，通常更适合高维协变量场景，但均需要通过交叉拟合来控制过拟合和偏差[39-40]。

2.3.4 案例

POUNDS Lost试验的事后分析使用多种机器学习方法估计不同基线特征下高脂与低脂饮食干预的CATE，并比较不同协变量集和学习器下基于CATE的治疗推荐表现[38]；Look AHEAD试验的事后分析也采用因果森林探索减重干预的HTE[41]。在重症医学中，因果森林、BART或效应评分方法也被用于评估多种干预措施在不同患者中的异质性反应，如氧疗目标与呼吸机策略、地塞米松剂量、肝素抗凝以及脓毒症抗生素时机[42-45]。研究结果提示，在临床综合征高度异质的领域，ATE接近无效并不意味着所有患者均无获益；相反，可能存在获益者和受害者相互抵消。

2.3.5 优势与局限

效应建模的优势是灵活、可整合多维患者特征、适合探索复杂HTE，并可为个体化治疗策略提供量化基础。其局限是对样本量、数据质量、治疗重叠、模型验证和临床解释要求更高。由于个体真实治疗效应不可观测，效应建模不能像普通预测模型那样直接验证预测值[46-47]。因此，模型输出应通过分层效应梯度、校准、外部验证、敏感性分析和临床可解释性综合评价。为便于比较，本文从主要问题、因果估计量、流行病学基础、适用场景、优势、局限和证据定位等方面总结了亚组分析、风险建模和效应建模三类HTE分析策略的差异，同时比较了模型特定和模型无关的优势和局限，见表4和表5。

表3 常见的元学习器

Table 3. Common methods for Meta-learners

方法	策略	适用场景
S-learner	单模型，治疗作为协变量	协变量少、治疗信号弱
T-learner	治疗组与对照组分别建模	两组结局函数可能不同
X-learner	用对照组信息改进治疗组估计（反之亦然）	两组样本量严重不平衡
DR-learner	基于双稳健伪结局	需要同时规避结局模型和倾向评分模型误设
R-learner	基于结局和倾向评分的残差正交化	高维设定、关注效应修饰本身

表4 HTE三类分析策略的特征比较

Table 4. Characteristics comparison of three types of HTE analysis strategies

维度	亚组分析	风险建模	效应建模
主要问题	某个预设亚组是否效应不同	哪些患者因基线风险高而绝对获益更大	多维患者特征如何共同影响治疗效应
因果估计量	亚组平均治疗效应	风险层内治疗效应	条件平均治疗效应
流行病学基础	效应修饰与交互作用	预后风险与绝对获益	多维效应修饰
适用场景	强先验、少数变量、监管或机制问题	总体治疗有效、风险差异明显、有可靠风险模型	大样本、高维变量、复杂非线性效应修饰
优势	简单、透明、临床可解释	稳健、临床决策意义强、易转化为NNT/净获益	灵活，可捕捉复杂交互，可支持个体化推荐
局限	多重比较、效能低、参考类问题	可能缺乏机制解释，可能遗漏非风险驱动HTE	易过拟合、解释困难、验证挑战大
证据定位	验证先验假设或生成假设	支持风险分层治疗决策	探索复杂HTE或构建治疗推荐模型

注：HTE.治疗效果异质性；NNT.治疗需要数。

表5 效应建模中模型特定与模型无关方法比较

Table 5. Comparison of model-specific and model-agnostic methods in effect modelling strategy

维度	模型特定方法	模型无关方法
基本思想	HTE的表达依赖特定模型结构或算法形式	先定义因果估计量，再选择任意模型估计组成部分
代表方法	回归交互、惩罚回归、因果树、因果森林、BART、BCF	标准化、IPW、AIPW、TMLE、S/T/X/DR/R-learner
优势	结构清晰，部分方法解释性强，可直接描述效应修饰模式	灵活性高，可结合机器学习，强调目标估计量而非固定模型
局限	依赖模型设定、变量编码和调参，可能过拟合	需要高质量辅助参数模型、交叉拟合和严格验证
适用问题	验证或解释某些效应修饰模式	估计目标人群或个体特征条件下的CATE

注：AIPW.增强逆概率加权；BART.贝叶斯加性回归树；BCF.贝叶斯因果森林；HTE.治疗效果异质性；IPW.逆概率加权；TMLE.目标最大似然估计；DR.双重稳健；CATE.条件平均治疗效应。

3 HTE研究的实施流程、可信度评价、报告与临床转化

3.1 实施流程

HTE研究应遵循“问题-设计-估计-验证”的流程。

第一，明确研究目标。HTE研究可分为四类目标：确认预设亚组效应、描述HTE程度、发现临床重要亚组、预测条件或个体化治疗效应。确认性研究需要强先验和预设分析；发现性研究应定位为假设生成；预测性研究需要额外验证和转化评估。

第二，完善研究设计。首先定义目标试验（target trial）。对于RWD研究，应明确目标人群、治疗策略、对照策略、零时点、随访起止、结局定义、因果对比和主要效应量表。所有候选效应修饰因素应在治疗前测量。接着明确变量角色，应区分混杂因素、预后因素、效应修饰因素、中介变量、碰撞变量等。混杂因素用于控制偏倚，预后因素用于风险建模，效应修饰因素用于定义或预测HTE，治疗后的变量不应作为基线效应修饰因素。为保证因果效应估计有效，对于RWD应优先使用倾向评分、加权、标准化或双稳健估计等方法控制混杂。如果总体研究设计存在严重偏倚，HTE结果不应被过度解释。

第三，估计并解释HTE。选择HTE建模方法。先验问题使用亚组分析；绝对获益分层问题使用风险建模；复杂多维效应修饰问题使用效应建模。若使用机器学习，应进行样本拆分、交叉验证、交叉拟合或外部验证。估计并解释治疗效应。HTE估计结果应优先报告绝对效应，同时报告相对效应。对于时间事件结局，应明确固定时间点RD、RR或受限平均生存时间差等效应测量指标。

第四，验证和转化。HTE结果需要通过内部验证、外部验证、敏感性分析和临床可解释性评价。若目标是进入临床决策支持系统，还应评估模型可用性、数据获取时点、更新机制、监管要求和实施障碍。

3.2 可信度评价

HTE结果的可信度评价不应仅依赖交互作用*P*值或单一模型指标，而应结合研究目标、效应量表、因果识别、统计可靠性、模型性能和临床可转化性进行综合判断。本文评价维度和表6条目主要参考ICEMAN效应修饰可信度评价工具、PATH声明及其解释说明文件，以及RWD中HTE研究的方法学框架[4, 8, 18]。

对于RCT中的HTE分析，应重点关注效应修饰变量是否预设、是否具有临床或生物学依据、交互方向和效应大小是否可信、是否存在多重比较问题，以及效应差异是否具有临床意义；对于RWD中的HTE分析，还应进一步关注混杂控制、治疗重叠、极端权重、有效样本量、未测量混杂和选择性删失等问题。

3.3 规范报告

HTE研究的规范报告条目主要依据PATH声明及其解释说明文件关于预测性HTE分析的报告建议，并结合ICEMAN工具、效应修饰/交互作用报告建议以及RWD中HTE研究的方法学要求综合整理[4, 8, 18]。报告内容应覆盖研究目标、研究设计、效应修饰变量、效应量表、亚组或风险层内结局风险、效应修饰指标、混杂控制、治疗重叠、模型验证和结果解释边界等方面。具体而

表6 HTE研究可信度评价清单

Table 6. Credibility evaluation checklist of HTE research

评价维度	关键问题
研究目标	分析目标是确认、描述、发现还是预测？证据解释强度是否与目标匹配？
先验性	效应修饰变量是否预设？是否有临床、药理或生物学依据？
效应测量指标	是否说明HTE在何种量表上评价？是否报告绝对效应和相对效应？
混杂控制	RWD研究是否通过设计和分析实现治疗组与对照组可比？
正性与重叠	各亚组或特征组合下是否存在足够治疗重叠？是否有极端权重？
多重比较	是否限制亚组数量或透明报告多重检验问题？
模型验证	是否进行交叉验证、外部验证、校准和敏感性分析？
临床意义	效应差异是否跨越临床决策阈值？是否改变NNT、NNH或净获益？
可转化性	亚组或效应评分是否临床可测量、可解释、可实施？

注：NNT.治疗需要数；NNH.伤害需要数。

言，HTE研究报告应至少包括以下内容：第一，明确研究目标属于确认、描述、发现还是预测。第二，说明研究设计是否基于目标试验仿真，明确零时点、治疗策略、对照策略和随访。第三，说明效应修饰变量是否治疗前测量，并解释其临床或机制依据。第四，报告各亚组或风险层内治疗组和对照组的结局风险，使读者能够理解绝对效应和相对效应。第五，报告效应修饰指标，例如RD之差或RR之比，而不只报告交互*P*值。第六，说明混杂控制、重叠性、极端权重和有效样本量。第七，若使用机器学习，应报告模型开发、调参、交叉验证、外部验证、校准、不确定性和临床可解释性。第八，明确区分验证性结论和探索性发现。

3.4 临床与监管转化

HTE研究的最终目标是服务决策，而不是生成复杂模型。临床转化需要回答四个问题：治疗效应差异是否足够大，足以改变治疗选择；识别出的高获益或高风险人群是否能够被临床稳定识别；治疗推荐是否同时考虑获益、伤害、患者偏好、成本和可及性；模型或规则是否经过外部验证和前瞻性评价。

在药品监管中，HTE证据可用于特殊人群安全性评价、风险管理计划、标签更新和适应证细化。在医保支付中，HTE证据可用于识别高绝对获益人群，支持分层支付或优先覆盖。在临床指南中，HTE证据可用于从“所有患者均推荐”转向“基于风险和治疗反应分层推荐”。但对于数据驱动发现的HTE，特别是机器学习预测的CATE，应避免直接作为强推荐依据，除非经过独立验证并具备临床可解释性。

4 结语

HTE分析的核心，是将药物评价从总体平均效应推进到患者特征条件下的因果效应评价。它既不是简单的事后亚组分析，也不是普通预测模型，而是建立在目标试验、因果识别、效应量表和临床决策问题基础上的方法学体系。

亚组分析适合验证少数具有强先验的效应修饰因素；风险建模适合评估基线风险驱动的绝对获益差异；效应建模适合在高维数据中探索和预测复杂HTE。RWD为HTE研究提供了更大的样本量、更广泛的患者异质性和更真实的临床场景，但也带来了更强的混杂、治疗选择机制和正性挑战。因此，RWD中的HTE研究应首先保证总体因果效应估计的有效性，再谨慎开展HTE分析。

未来药物流行病学HTE研究应从“寻找显著亚组”转向“生成可解释、可验证、可行动的治疗效应差异证据”，并在报告中明确区分验证性分析和探索性分析、绝对效应和相对效应、预后风险和治疗反应、模型预测和因果效应。只有这样，HTE方法才能真正服务于药物有效性、安全性、获益-风险评价和精准用药决策。

利益冲突声明：作者声明本文不存在任何经济或非经济利益冲突。

参考文献

1 Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification[J]. JAMA, 2007, 298(10): 1209-1212. DOI: 10.1001/jama.298.10.1209.

2 Kent DM, Rothwell PM, Ioannidis JP, et al. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal[J]. Trials, 2010, 11: 85. DOI: 10.1186/1745-6215-11-85.

3 Varadhan R, Segal JB, Boyd CM, et al. A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research[J]. J Clin Epidemiol, 2013, 66(8): 818-825. DOI: 10.1016/j.jclinepi.2013.02.009.

4 Segal JB, Varadhan R, Groenwold RHH, et al. Assessing heterogeneity of treatment effect in real-world data[J]. Ann Intern Med, 2023, 176(4): 536-544. DOI: 10.7326/M22-1510.

5 Thelen H, Hennessy S. Characterizing treatment effect heterogeneity using real-world data[J]. Clin Pharmacol Ther, 2025, 117(5): 1209-1216. DOI: 10.1002/cpt.3627.

6 Rubin DB. Causal inference using potential outcomes: design, modeling, decisions[J]. J Am Stat Assoc, 2005, 100(469): 322-331. DOI: 10.1198/016214504000001880.

7 Hernan MA, Robins JM. Causal inference: What if (First edition)[B]. Taylor and Francis, 2024.

8 Kent DM, Paulus JK, van Klaveren D, et al. The predictive approaches to treatment effect heterogeneity (PATH) statement[J]. Ann Intern Med, 2020, 172(1): 35-45. DOI: 10.7326/M18-3667.

9 VanderWeele TJ, Knol MJ. A tutorial on interaction[J]. Epidemiol Methods, 2014, 3(1): 33-72. https://hsph.harvard.edu/wp-content/uploads/2024/10/InteractionTutorial_EM-1.pdf

10 Rubin D. Estimating causal effects of treatments in experimental and observational studies[J]. ETS Res Bull Ser, 1972, 1972(2): 688-701. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/j.2333-8504.1972.tb00631.x

11 Knol MJ, VanderWeele TJ. Recommendations for presenting

analyses of effect modification and interaction[J]. Int J Epidemiol, 2012, 41(2): 514-520. DOI: 10.1093/ije/dyr218.

12 Hernan 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.

13 Hernan MA, Sauer BC, Hernandez-Diaz S, et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses[J]. J Clin Epidemiol, 2016, 79: 70-75. DOI: 10.1016/j.jclinepi.2016.04.014.

14 Hernan MA, Dahabreh IJ, Dickerman BA, et al. The target trial framework for causal inference from observational data: Why and when is it helpful?[J]. Ann Intern Med, 2025, 178(3): 402-407. DOI: 10.7326/ANNALS-24-01871.

15 Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: Importance, indications, and interpretation[J]. Lancet, 2005, 365(9454): 176-186. DOI: 10.1016/S0140-6736(05)17709-5.

16 Assmann SF, Pocock SJ, Enos LE, et al. Subgroup analysis and other (mis) uses of baseline data in clinical trials[J]. Lancet, 2000, 355(9209): 1064-1069. DOI: 10.1016/S0140-6736(00)02039-0.

17 Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine--reporting of subgroup analyses in clinical trials[J]. N Engl J Med, 2007, 357(21): 2189-2194. DOI: 10.1056/NEJMsr077003.

18 Schandelmaier S, Briel M, Varadhan R, et al. Development of the instrument to assess the credibility of effect modification analyses (iceman) in randomized controlled trials and meta-analyses[J]. CMAJ, 2020, 192(32): E901-E906. DOI: 10.1503/cmaj.200077.

19 Maemondo M, Inoue A, Kobayashi K, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR[J]. N Engl J Med, 2010, 362(25): 2380-2388. DOI: 10.1056/NEJMoa0909530.

20 Kim ES, Hirsh V, Mok T, et al. Gefitinib versus docetaxel in previously treated non-small-cell lung cancer (interest): a randomised phase iii trial[J]. Lancet, 2008, 372(9652): 1809-1818. DOI: 10.1016/S0140-6736(08)61758-4.

21 Thatcher N, Chang A, Parikh P, et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: Results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation In Lung Cancer)[J]. Lancet, 2005, 366(9496): 1527-1537. DOI: 10.1016/S0140-6736(05)67625-8.

22 Fox KA, Poole-Wilson P, Clayton TC, et al. 5-year outcome of an interventional strategy in non-ST-elevation acute coronary syndrome: the british heart foundation rita 3 randomised trial[J]. Lancet, 2005, 366(9489): 914-920. DOI: 10.1016/S0140-6736(05)67222-4.

23 Tang H, Donahoo WT, Svensson M, et al. Heterogeneous treatment effects of sodium-glucose cotransporter 2 inhibitors on risk of dementia in people with type 2 diabetes: a population-based cohort study[J]. Alzheimers Dement, 2024, 20(8): 5528-5539. DOI: 10.1002/alz.14048.

24 Wang T, Keil AP, Buse JB, et al. Glucagon-like peptide 1 receptor agonists and asthma exacerbations: which patients benefit most?[J]. Ann Am Thorac Soc, 2024, 21(11): 1496-1506. DOI: 10.1513/AnnalsATS.202309-836OC.

25 Rekkas A, Rijnbeek PR, Kent DM, et al. Estimating individualized treatment effects from randomized controlled trials: a simulation study to compare risk-based approaches[J]. BMC Med Res Methodol, 2023, 23(1): 74. DOI: 10.1186/s12874-023-01889-6.

26 van Klaveren D, Balan TA, Steyerberg EW, et al. Models with interactions overestimated heterogeneity of treatment effects and were prone to treatment mistargeting[J]. J Clin Epidemiol, 2019, 114: 72-83. DOI: 10.1016/j.jclinepi.2019.05.029.

27 Selby JV, Maas C, Fireman BH, et al. Predictive modeling of heterogeneous treatment effects in rcts: a scoping review[J]. JAMA Netw Open, 2025, 8(7): e2522390. DOI: 10.1001/jamanetworkopen.2025.22390.

28 Feuerriegel S, Frauen D, Melnychuk V, et al. Causal machine learning for predicting treatment outcomes[J]. Nat Med, 2024, 30(4): 958-968. DOI: 10.1038/s41591-024-02902-1.

29 Abecassis J, Dumas E, Alberge J, et al. From prediction to prescription: machine learning and causal inference for the heterogeneous treatment effect[J]. Annu Rev Biomed Data Sci, 2025, 8(1): 381-404. DOI: 10.1146/annurev-biodatasci-103123-095750.

30 Inoue K, Adomi M, Efthimiou O, et al. Machine learning approaches to evaluate heterogeneous treatment effects in randomized controlled trials: a scoping review[J]. J Clin Epidemiol, 2024, 176: 111538. DOI: 10.1016/j.jclinepi.2024.111538.

31 Lipkovich I, Svensson D, Ratitch B, et al. Modern approaches for evaluating treatment effect heterogeneity from clinical trials and observational data[J]. Stat Med, 2024, 43(22): 4388-4436. DOI: 10.1002/sim.10167.

32 Imai K, Ratkovic M. Estimating treatment effect heterogeneity in randomized program evaluation[J]. Ann Appl Stat, 2013, 7(1): 443-470. DOI: 10.1214/12-Aoas593.

33 Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects[J]. Proc Natl Acad Sci U S A, 2016, 113(27): 7353-7360. DOI: 10.1073/pnas.1510489113.

34 Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests[J]. J Am Stat Assoc, 2018, 113(523): 1228-1242. DOI: 10.1080/01621459.2017.1319839.

35 Hill JL. Bayesian nonparametric modeling for causal inference[J]. J Comput Graph Stat, 2011, 20(1): 217-240. DOI: 10.1198/jcgs.2010.08162.

36 Hahn PR, Murray JS, Carvalho CM. Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects (with discussion) [J]. Bayesian Analysis, 2020, 15(3): 965-1056.

37 Kunzel SR, Sekhon JS, Bickel PJ, et al. Metalearners for estimating heterogeneous treatment effects using machine learning[J]. Proc Natl Acad Sci U S A, 2019, 116(10): 4156-4165. DOI: 10.1073/pnas.1804597116.

38 Hamaya R, Hara K, Manson JE, et al. Machine-learning

approaches to predict individualized treatment effect using a randomized controlled trial[J]. Eur J Epidemiol, 2025, 40(2): 151-166. DOI: 10.1007/s10654-024-01185-7.

39 Nie X, Wager S. Quasi-oracle estimation of heterogeneous treatment effects[J]. Biometrika, 2021, 108(2): 299-319. DOI: 10.1093/biomet/asaa076.

40 Kennedy EH. Towards optimal doubly robust estimation of heterogeneous causal effects[J]. Electro J Stat, 2023, 17(2): 3008-3049. DOI: 10.1214/23-EJS2157.

41 Baum A, Scarpa J, Bruzelius E, et al. Targeting weight loss interventions to reduce cardiovascular complications of type 2 diabetes: a machine learning-based post-hoc analysis of heterogeneous treatment effects in the look ahead trial[J]. Lancet Diabetes Endocrinol, 2017, 5(10): 808-815. DOI: 10.1016/S2213-8587(17)30176-6.

42 Munroe ES, Spicer A, Castellvi-Font A, et al. Evidence-based personalised medicine in critical care: a framework for quantifying and applying individualised treatment effects in patients who are critically ill[J]. Lancet Respir Med, 2025, 13(6): 556-568. DOI: 10.1016/S2213-2600(25)00054-2.

43 Blette BS, Granholm A, Li F, et al. Causal bayesian machine learning to assess treatment effect heterogeneity by dexamethasone dose for patients with COVID-19 and severe hypoxemia[J]. Sci Rep, 2023, 13(1): 6570. DOI: 10.1038/s41598-023-33425-3.

44 Goligher EC, Lawler PR, Jensen TP, et al. Heterogeneous treatment effects of therapeutic-dose heparin in patients hospitalized for COVID-19[J]. JAMA, 2023, 329(13): 1066-1077. DOI: 10.1001/jama.2023.3651.

45 Hechtman RK, Kipnis P, Cano J, et al. Heterogeneity of benefit from earlier time-to-antibiotics for sepsis[J]. Am J Respir Crit Care Med, 2024, 209(7): 852-860. DOI: 10.1164/rccm.202310-1800OC.

46 Desai RJ, Glynn RJ, Solomon SD, et al. Individualized treatment effect prediction with machine learning-salient considerations[J]. NEJM Evid, 2024, 3(4): EVIDoa2300041. DOI: 10.1056/EVIDoa2300041.

47 Moller M, Wild EM, Tan W, et al. Estimating heterogeneous treatment effects with real-world health data: a scoping review of machine learning methods[J]. Value Health, 2026, 29(5): 905-913. DOI: 10.1016/j.jval.2026.01.013.

收稿日期：2026 年 06 月 13 日 修回日期：2026 年 06 月 23 日
本文编辑：李绪辉 桂裕亮