

· 专家论坛 ·

DOI: 10.12449/JCH251105

人工智能相关肝病临床试验的方法学考量

史立臣^a, 冯睿时^a, 贾继东^b, 尤红^b, 孔媛媛^a

首都医科大学附属北京友谊医院 a. 北京市临床医学研究所临床流行病学与循证医学研究室, b. 肝病中心, 北京 100050

通信作者: 孔媛媛, konggy@ccmu.edu.cn (ORCID: 0000-0002-2586-1443)

摘要: 近年来,人工智能(AI)技术在肝病临床试验领域的探索与发展,推动了肝病临床试验的不断创新。AI已逐渐成为患者筛选、风险预测、终点评价及结果解读等临床试验环节的重要技术工具。然而,AI在临床试验中的规范化应用仍面临数据质控、模型可解释性及因果推断等方面的方法学挑战。本文基于方法学视角,针对AI作为研究对象(验证型)和AI作为研究工具(辅助型)两类临床试验,系统梳理其在肝病领域的主要应用场景,探讨相关方法学挑战及解决策略,旨在为促进该技术的规范化应用提供方法学参考。

关键词: 肝疾病; 临床试验; 人工智能; 方法

Application of artificial intelligence in clinical trials of liver diseases: A methodological perspective

SHI Lichen^a, FENG Ruishi^a, JIA Jidong^b, YOU Hong^b, KONG Yuanyuan^a

a. Clinical Epidemiology and EBM Unit of Beijing Clinical Research Institute, b. Liver Research Center, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China

Corresponding author: KONG Yuanyuan, konggy@ccmu.edu.cn (ORCID: 0000-0002-2586-1443)

Abstract: In recent years, the exploration and development of artificial intelligence (AI) technology in clinical trials for liver diseases have promoted the continuous innovation of research methods and processes in this field. AI has gradually become an important technical tool for various links of clinical trial including patient selection, risk stratification, endpoint evaluation, and result interpretation. Nevertheless, the standardized integration of AI into clinical trials still faces the methodological challenges such as data quality control, model interpretability, and causal inference. From the perspective of methodology, this article systematically reviews the principal application scenarios of AI as an object under investigation (validation trials) and as a research tool (supportive trials) in clinical trials for liver diseases, as well as the major methodological challenges of AI-related clinical trials along and the corresponding solution strategies, in order to provide methodological guidance for promoting the scientific and standardized implementation of AI technologies.

Key words: Liver Diseases; Clinical Trial; Artificial Intelligence; Methods

近年来,人工智能(artificial intelligence, AI)在医学领域,包括肝病临床试验领域迅速发展。在大数据和精准医疗的推动下,AI技术已逐渐成为优化患者筛选、风险预测、终点评价及结果解读等关键环节的重要工具,为创新临床试验设计、优化实施流程带来了新机遇,成为驱动肝病临床试验向高质量和智能化方向转型的重要技术动力。AI相关临床试验主要分为两种类型:一类是AI作为研究对象,即将AI系统本身视作医疗器械或临床技术,通过临

床试验系统评估其在诊断、预测或决策支持中的安全性与有效性,本文将其归类为“AI验证型临床试验”;另一类是AI作为研究工具,并非试验的主要干预对象,而是在试验设计与实施过程中辅助完成患者筛选、风险分层、结局评价或数据管理等任务,本文将其归类为“AI辅助型临床试验”。前者强调对AI自身临床价值的验证,后者则突出其在提升试验效率与质量方面的应用价值。

然而,AI在临床试验中的规范化应用仍面临诸多方

法学挑战,涉及数据质控、模型可解释性及因果推断能力等。本文将聚焦于肝病领域的相关应用场景,从方法学角度系统梳理AI在肝病临床试验中的应用现状和主要挑战,并结合国际主流AI方法学和医学研究报告规范,探讨相应的解决策略,旨在为促进AI赋能肝病临床试验的规范化应用与可持续发展提供科学参考。

1 AI相关肝病临床试验的应用与设计

1.1 AI相关肝病临床试验的基本情况 通过PubMed和ClinicalTrials.gov数据库检索至2025年7月21日已发表的文献及注册的临床试验,检索语言限于英文和中文,检索与肝脏疾病相关且涉及AI在肝病诊断、预后评估、疗效预测或治疗决策中的应用研究。初步检索到83篇文

献和236项注册临床试验,经筛选后最终纳入36项AI相关肝病临床试验,其中包含13篇文献和23项注册临床试验。

在纳入的36项AI相关临床试验中,31项为AI验证型临床试验,5项为AI辅助型临床试验。疾病分布以肝癌(61.1%)和脂肪肝(27.7%)为主。AI的应用领域涵盖6个方面:医学影像分析、数字病理分析、基于生物标志物的疾病分型与诊断辅助、临床决策支持、药物剂量个体化管理及实施策略优化等。总体上,AI验证型临床试验主要聚焦于影像分析和诊断支持,AI辅助型临床试验则侧重个体化药物管理和多模态数据融合。部分AI工具已获得监管机构批准,但多数处于试验阶段或尚未正式应用(表1)。

表1 现有AI验证型或AI辅助型肝病临床试验的基本情况
Table 1 Basic characteristics of existing AI-validated or AI-assisted clinical trials in liver diseases

试验注册号 (注册时间)	参与国家	疾病(病因)	AI名称	AI注册/批准	AI应用场景	出版物/试验状态
AI验证型临床试验						
NCT04722120 (2016-06)	韩国	HCC(HBV)	ClariACE	—	CT影像分析	Kang等 ^[1]
NCT03448757 (2018-02)	巴西	HCC	Artificial Intelligence program	—	生物信号分析	招募中
NCT06526754 (2019-01)	中国	HCC	AI-assisted automatic segmentation	—	医学影像三维重建	完成
NCT03917017 (2019-01)	中国	HCC	Radiomics and Watson artificial intelligence	—	临床决策支持+术前 风险评估+影像导航	未知
NCT04180631 (2019-03)	韩国	NAFLD	—	—	超声影像后处理	完成
NCT04027556 (2019-06)	韩国	HCC(HBV/HCV)	ClariCT.AI	FDA 510(k) K183460(2019)	CT影像分析	Bae等 ^[2]
NCT04024514 (2019-12)	韩国	HCC(HBV/HCV)	ClariACE	—	CT影像分析	Yoon等 ^[3]
NCT04083378 (2020-01)	美国	原发性或继发性 肝肿瘤	Raystation	FDA 510(k) K220141(2022)	CT影像后处理	Odisio等 ^[4]
NCT04462562 (2020-07)	韩国	NAFLD	2D CNN	—	超声影像分析	Jeon等 ^[5]
NCT04463667 (2020-08)	中国	MASLD	Siemens Liver lab	—	MRI影像后处理	Wu等 ^[6]
CTRI/2020/08/ 027072(2020-10)	印度	MAFLD	WBDT	—	AI辅助个体化管理	Joshi等 ^[7]
TCTR20201230003 (2020-12)	泰国	FLL(HCC/肝囊 肿/血管瘤)	AI assistance	—	超声影像分析	Tiyarattanachai等 ^[8]
NCT05804799 (2021-01)	韩国、德国	原发性或继发性 肝肿瘤	ClariCT.AI	FDA 510(k) K183460(2019)	CT影像分析	Lee等 ^[9]
NCT04579874 (2021-01)	美国	MASLD	LIVERFASt	—	基于血液生物标志 物的疾病分型与诊 断辅助	完成
NCT04843176 (2021-03)	中国	HCC	Prototype AI algorithm	—	CT影像分析	招募中

表 1(续)

Table 1 (continued)

试验注册号 (注册时间)	参与国家	疾病(病因)	AI名称	AI注册/批准	AI应用场景	出版物/试验状态
NCT04802954 (2021-09)	法国	HCC	—	—	超声影像分析	完成
NCT05182099 (2022-01)	韩国	FLL	AIRTM	—	医学影像重建	未知
NCT05676177 (2023-01)	以色列	NAFLD	—	—	CT影像后处理	招募中
NCT05782283 (2023-04)	美国	肝纤维化	ACE 2.0 model	—	心电图信号分析	招募完成
NCT06626087 (2023-11)	中国	HCC	Prototype AI algorithm	—	CT影像分析	招募中
NCT05681949 (2023-11)	德国	肝胆系统肿瘤	ADBoard	—	临床决策支持	未知
NCT06735118 (2023-12)	中国	脂肪肝(癌症患 者化疗前后)	—	—	MRI影像后处理	招募中
NCT06061328 (2024-02)	美国	HCC	OASIS决策辅助 工具	—	实施策略选择优化	未知
NCT06317181 (2024-04)	德国	弥漫性肝病和 FLL	—	—	超声影像后处理	招募中
NCT06463444 (2024-06)	中国	HCC	多模态深度学习 模型	—	MRI影像分析+数字 病理分析+基因检 测数据融合	招募中
NCT06334965 (2024-07)	法国	HCC	基于PET-MRI的 多模态影像AI	—	MRI影像分析	招募中
NCT06888310 (2024-12)	意大利	cACLD(NAFLD)	AI模型融合组学 数据	—	多模态数据融合	招募中
NCT06859840 (2025-03)	中国	HCC	LIDAR model	—	CT影像分析	尚未招募
NCT06944353 (2025-09)	美国	MASLD	STIRRED CDSS	—	自然语言处理,电子 健康记录分析	尚未招募
无(不详)	中国	HCC	M-RCNN	—	超声影像分析	Zhang等 ^[10]
无(不详)	韩国	HCC	POEM CDSS	—	AI辅助个体化管理	Kang等 ^[11]
AI辅助型临床试验						
NCT03449446 (2018-03)	美国、加拿大、 澳大利亚、中国、 新西兰	肝纤维化/肝硬 化(NASH)	PathAI	FDA 510(k) K212361(2022)	数字病理分析	Loomba等 ^[12]
NCT06311916 (2024-05)	中国	HCC	New-adj-Net	—	MRI影像分析+数字 病理分析+基因检 测数据融合	尚未招募
NCT06420440 (2024-06)	中国	HCC	New-adj-Net	—	MRI影像分析+数字 病理分析+基因检 测数据融合	招募中
NCT05669339 (2024-12)	美国	HCC	PPM	—	药物剂量个体化 调整	暂停(期中分析)
无(不详)	美国	PBC	AIMS	—	个体化药物剂量 管理	McMichael等 ^[13]

注:“—”表示无相关数据。cACLD,代偿期晚期慢性肝病;FDA,美国食品药品监督管理局;HCC,肝细胞癌;MAFLD/MASLD,代谢相关脂肪性肝病/代谢功能障碍相关脂肪性肝病;NAFLD,非酒精性脂肪性肝病;NASH,非酒精性脂肪性肝炎;PET,正电子发射断层显像术;PBC,原发性胆汁性胆管炎;WBDT,全身数字孪生;FLL,肝脏局灶性病变。

1.2 AI相关肝病临床试验的设计要点 纳入的AI相关 区组随机(7项),另有顺序分配、阶梯式楔形集群随机、肝病临床试验在设计方面呈现多样性。在分配方法方面, Pocock-Simon 动态最小化随机和整群随机各1项,提示以非随机分配最为常见(12项),其次为简单随机(9项)、 简便、可操作性强的随机化方式仍占主导地位。在盲法

设置方面,开放标签18项(50.0%),盲法16项(44.4%)。实际样本量范围为20~504例。在干预模式方面,试验组普遍引入AI技术或AI辅助优化手段,具体包括:单纯AI干预10项、AI联合医学影像优化或操作优化11项、AI输出与专家判读对比3项,以及AI辅助临床决策或联合新药/新方案干预12项;对照组均未涉及AI相关干预。所采用的AI算法包括深度学习、卷积神经网络、随机森林、机器学习、自然语言处理及多模态融合模型等。仅有2项研

究报告完全符合CONSORT^[11]或CONSORT-AI^[8]标准,其余研究仅部分满足相关报告要求,特别在统计学设计、AI算法细节及结果呈现等方面信息披露不足,提示AI相关肝病临床试验在报告质量和规范性方面有待提高(表2)。

2 AI相关临床试验的主要方法学挑战

2.1 数据集质量参差与共享障碍 当前AI在肝病临床试验领域的应用面临来自真实世界诊疗数据的多重挑战,包括数据的多源异构、质量参差及共享障碍等。不同医疗

表2 AI验证型或AI辅助型肝病临床试验的设计要点
Table 2 Key design elements of AI-validated or AI-assisted clinical trials in liver diseases

试验类型及试验注册号	随机化方法	盲法	样本量 (预期/实际)	试验组	对照组	AI算法
AI验证型临床试验						
NCT04722120(2016-06)	区组随机	双盲	68/68	低剂量CT+IMR重建+DL-CB算法后处理	标准剂量CT+IMR重建	CNN
NCT03448757(2018-02)	非随机	开放标签	60/—	HCC患者	健康志愿者	—
NCT06526754(2019-01)	简单随机	四盲	—/64	基于实体3D打印模型进行术前规划	基于AI辅助分割+人工修正的数字模拟进行术前规划	深度学习
NCT03917017(2019-01)	单臂试验	开放标签	100/—	放射组学+IBM Watson AI辅助决策组	无	—
NCT04180631(2019-03)	单臂试验	开放标签	—/124	使用QUS检测肝脂肪	无	深度学习
NCT04027556(2019-06)	区组随机	双盲	90/90	基于LBW的低对比剂量	标准对比剂量	DLD
NCT04024514(2019-12)	区组随机	双盲	52/52	低碘量(300 mgI/kg)+深度学习图像增强	常规碘量(525 mgI/kg)+混合迭代重建	CNN
NCT04083378(2020-01)	Pocock-Simon动态最小化随机	单盲	107/50	使用基于AI的辅助软件对消融边界进行定量评估	传统视觉方法手动评估消融边界	—
NCT04462562(2020-07)	单臂试验	开放标签	173/173	2D CNN深度学习模型计算USFF	MRI PDFF	2D CNN
NCT04463667(2020-08)	—	—	100/80	autoWLFatQ	MRS和avePDFF	自动肝脏分割算法
CTRI/2020/08/027072(2020-10)	区组随机	开放标签	700/319	数字治疗组	标准护理组	GBDT, DLNN, E-SLSTM
TCTR20201230003(2020-12)	交叉随机	双盲	309/504	使用AI辅助进行超声检查	不使用AI辅助进行超声检查	CNN
NCT05804799(2021-01)	—	单盲	300/296	使用单管B(33%标准剂量)扫描数据,通过ClariCT.AI进行图像重建	使用双源CT扫描(管A+B,标准剂量),图像通过高级模型迭代重建	DLD
NCT04579874(2021-01)	单臂试验	开放标签	—/241	LIVERFASt	无	—
NCT04843176(2021-03)	简单随机	单盲	250/—	Prototype AI algorithm	由2位腹部放射科专家依据LI-RADS标准独立评估	Prototype AI
NCT04802954(2021-09)	非随机	开放标签	—/400	肝细胞癌结节>1 cm的患者,视频采集	无肝癌患者,随访1年确认无新结节,视频采集	CNN

表 2(续)

Table 2 (continued)

试验类型及试验注册号	随机化方法	盲法	样本量 (预期/实际)	试验组	对照组	AI算法
NCT05182099(2022-01)	非随机	单盲	52/—	使用 AIRTm 对 Gd-EOB-DTPA 增强肝脏 MRI 图像进行重建	使用传统的图像重建算法对 Gd-EOB-DTPA 增强肝脏 MRI 图像进行重建	深度学习
NCT05676177(2023-01)	单臂试验	开放标签	150/—	采用超低剂量 CT 和 MRI 同时扫描的受试者	无	深度学习
NCT05782283(2023-04)	群组随机化	开放标签	400/—	使用 ACE 2.0 模型进行早期肝硬化检测并提醒医生	常规护理组	CNN
NCT06626087(2023-11)	简单随机	单盲	250/—	Prototype AI algorithm	由至少 10 年经验的腹部影像专科医师依据 LI-RADS 标准评估	—
NCT05681949(2023-11)	简单随机	单盲	1 200/—	肿瘤委员会使用 ADBoard	肿瘤委员会不使用 ADBoard	—
NCT06735118(2023-12)	简单随机化	双盲	120/—	接受化疗癌症患者	不接受化疗癌症患者	深度学习
NCT06061328(2024-02)	整群随机	开放标签	8 020/—	基于机器学习的 OASIS 决策辅助工具的医疗中心	基于专家意见的现有决策辅助工具的医疗中心	机器学习
NCT06317181(2024-04)	非随机	开放标签	200/—	超声射频数据采集	弹性成像数据采集	深度学习
NCT06463444(2024-06)	单臂试验	开放标签	30/—	HAIC+替雷利珠单抗+仑伐替尼联合治疗	无	多模态深度学习
NCT06334965(2024-07)	单臂试验	开放标签	180/—	PET-MRI 多模态成像	无	—
NCT06888310(2024-12)	单臂试验	开放标签	408/—	MASLD 且肝硬度测定 LSM \geq 10 kPa 的患者, 进行额外血液采样	无	—
NCT06859840(2025-03)	简单随机	单盲	10 000/—	LIDAR AI 辅助诊断	常规影像诊断	—
NCT06944353(2025-09)	阶梯式楔形集群随机	开放标签	4 704/—	使用 CDSS 在急诊科提示医生通知患者并提供随访推荐	常规护理, 无特定干预	机器学习+自然语言处理
无(不详)	简单随机	双盲	—/100	多模态组(联合彩色多普勒超声检查+超声造影+超声弹性成像检查进行诊断)	彩色多普勒超声检查进行诊断; 超声造影进行诊断; 超声弹性成像检查进行诊断	M-RCNN
无(不详)	—	双盲	38/40	护士接受 POEM CDSS 系统+自学手册	护士仅接受自学手册	RF
AI 辅助型临床试验						多模态深度学习
NCT03449446(2018-03)	群组随机	双盲	350/392	selonsertib、cilofexor、firsocostat 单药治疗及两两联合	安慰剂	CNN
NCT06311916(2024-05)	简单随机	开放标签	312/—	HAIC+替雷利珠单抗+仑伐替尼+肝切除	直接肝切除	AI 辅助型临床试验
NCT06420440(2024-06)	简单随机	开放标签	160/—	术前: HAIC+替雷利珠单抗+仑伐替尼; 术后: 替雷利珠单抗	术后: 替雷利珠单抗	多模态深度学习
NCT05669339(2024-12)	顺序分配	开放标签	12/—	索拉非尼、索尼得吉、伊立替康 3 种药物组合, 按 3+3 剂量递增设计分配剂量	无	—
无(不详)	群组随机	—	—/20	中、高浓度他克莫司	低浓度他克莫司	—

注: “—”表示无相关数据。autoWLFatQ, 自动化全肝脂肪定量; avePDFF, 平均质子密度脂肪分数; CDSS, 临床决策支持系统; CNN, 卷积神经网络; DL-CB, 基于深度学习的碘对比增强算法; DLD, 基于深度学习的去噪算法; DLNN, 深度学习神经网络; E-SLSTM, 扩展序列长短期记忆网络; GBDT, 梯度提升决策树; HAIC, 肝动脉灌注化疗; LBW, 去脂体质量; LI-RADS, 肝脏影像报告和数据系统; LSM, 肝硬度值; MASLD, 代谢相关脂肪性肝病; M-RCNN, 基于掩模区域的卷积神经网络分割算法; OASIS, 优化实施策略选择的方法; PDFF, 质子密度脂肪分数; POEM CDSS, 栓塞后综合征管理临床决策支持系统; QUS, 定量超声; USFF, 超声脂肪分数; IMR, 迭代模型重建; MRS, 磁共振波谱成像; RF, 随机森林。

机构间的数据常存在格式不统一、编码体系各异和数据缺失等现象,且行业内缺乏统一的、覆盖多病种和全诊疗流程的数据标准。尽管国际上已有临床数据交换标准协会等制定注册临床试验数据规范,但此类标准尚难直接应用于多源、多模态、和动态变化的真实世界临床数据。

为突破数据壁垒,实现多源异构数据互联互通,行业正逐步推广与应用多项高新技术,包括建立标准化数据元体系与医学术语库,实现数据结构和内容的统一匹配;采用数据标识与语义对齐技术(如自动化字段映射引擎),提升不同系统间的数据互操作性;引入数据自动治理智能体,通过规则和机器学习实时提升数据完整性和一致性;同时加强数据安全与隐私保护,融合密态计算、联邦学习和可信执行环境,保障跨机构间数据在安全前提下的可信共享。

2.2 模型偏倚与泛化能力不足 在肝病风险预测AI模型的开发过程中,模型偏倚与泛化能力始终是核心挑战。训练与验证过程中的选择偏倚是影响建模有效性和可推广性的关键问题之一。据文献报道,目前大多数研究模型的训练以及验证主要依赖于单中心或区域性研究,样本量从几十例到十余万例不等,其中大部分研究属于小样本队列,仅约30%的模型进行了独立的外部数据验证,大部分仍采用“训练集/测试集”按7:3或8:2比例划分内部数据进行验证。这直接削弱了模型的泛化能力和临床适用性^[14]。

随着真实世界医疗数据的不断积累,通过跨中心、跨区域的数据共享和验证,以及多时间段、多种族的数据整合,有望进一步增强模型的稳定性及泛化能力^[15]。未来模型研发,除强调在原始训练数据上的性能外,外部独立验证应成为常规流程,纳入多样本、多中心长期队列数据是提升模型可信度的关键。

2.3 模型透明度与可解释性差 传统AI模型特别是深度学习(如神经网络)往往被视为“黑箱”,其内部决策机制不透明、难以解释。一项肝病AI模型的系统综述显示,当前AI相关的肝病研究主要采用随机森林和线性模型等可解释性方法,这类模型使用较为广泛,但根据影响力评分 $[s = \text{AUROC} \times \log_{10}(n)]$, where $n = \text{validation cohort size}$ (scaled to 1.0 ~ 1.5);基于模型AUROC与验证队列规模加权,其实际影响力相对较低^[16]。近年来,深度学习等方法在肝病多模态数据(如组织病理学、组学)中的应用日益增加,展现出处理高维复杂数据的处理优势,且在大型队列和复杂疾病(如HCC、慢加急性肝衰竭)的建模中表现出更高的准确性和影响力。此外,目前大多数研究仍采用

特征工程与建模过程分离的策略,端到端建模尚不普遍。综上,肝病AI建模领域正在由传统、可解释模型逐步向深度学习等复杂模型的多元化方向发展。

以AI辅助病理诊断为例,可解释AI方法不仅能够输出肿瘤与非肿瘤的预测结果,还能清晰展示模型决策背后的关键组成部分。首先,透明度指的是让医生了解模型内部的神经网络结构和每一步神经元的激活模式,揭示模型的“思考过程”;其次,语义体现为明确识别那些促使模型做出判断的图像特征,例如哪些细胞核形态或组织结构引起了模型的关注并影响最终结果;最后,解释是在整体上揭示这些关键特征和网络响应是如何共同作用、最终使得模型给出肿瘤或非肿瘤结果。这三方面共同提升了AI模型的可解释性和可信度,让医生能够结合自身专业知识,更好地理解和甄别模型辅助诊断结果^[17]。

2.4 AI对传统试验设计与统计分析的挑战 AI的应用为医学研究创造了新的可能,同时也给传统试验设计带来了挑战。AI模型通常需要高维度、大规模、动态变化的数据输入,这与传统临床试验依赖的严格分组、精确样本量估算及随机对照设计存在本质区别。AI研究多采用回顾性数据或真实世界数据,研究人群具有较高异质性,干预和暴露因素难以精确控制,导致经典随机对照试验的核心原则(随机化分配、盲法实施和对对照组设定等)难以直接应用。为解决这一问题,研究者需要参考AI相关的临床试验设计规范,以确保研究设计的科学性和临床推广价值。

同时,AI方法也对传统统计分析框架提出了新要求。AI算法主要关注数据变量间的统计相关性和非线性模式识别,而非直接揭示因果关系,这使得模型可能在具备较高预测准确性的同时,缺乏对疾病机制、混杂因素和干预效应的准确判断能力。在评估医学干预措施和研究终点时,忽视因果推断可能导致模型结果包含系统性偏倚,降低研究结论的可解释性和临床实用性。针对这些问题,国际学术界已提出多项指导原则和标准(表3),以及整合因果推断理论的创新分析方法^[18],从而有助于提升AI模型的科学性和可解释性。

2.5 伦理与合规的挑战 除技术与方法学问题,AI相关临床试验亦面临伦理与合规挑战。首先,知情同意应明确涵盖AI工具的功能、潜在风险及其局限性,确保临床试验参与者在充分理解后自愿参与。其次,数据隐私与安全问题需严格把控,AI模型训练通常依赖多中心敏感数据,其跨机构共享与应用应严格遵守合规要求,防止隐私泄露和数据滥用。再次,需关注算法公平性,避

免因数据分布不均或样本代表性不足导致对特定人群的系统性偏差。最后,责任归属尚不明确,当AI辅助诊疗发生误判时,研究者、开发者及监管机构之间的责任界定仍存在模糊地带。总体而言,AI临床试验的伦理与合规规范尚不完善,这在一定程度上制约了其规范化和可持续发展。为此,国际上已发布多项规范为相关实践提供参考(表3)。

3 AI相关临床研究设计与报告规范

随着AI在医学领域的应用日益广泛,国际学术界相继提出了针对AI临床研究的规范,以弥补传统临床试验设计和报告在AI场景下的不足。研究表明,尽管当前指南已对数据质量控制、模型透明性、外部验证和效果评价等关键环节提出了明确要求,但在实际执行过程中仍面临诸多挑战,包括多中心数据标准不统一、模型验证不足、指标报告单一以及对AI模型动态迭代过程缺乏监管等^[19]。为此,国际上已发布或正在制定的多项规范覆盖了AI医学研究的全流程(表4)。遵循相关规范,有助于提升肝病AI研究的透明度、可重复性和标准化水平,从而增强AI在肝病领域的临床转化与推广价值。

4 未来前景与方法学创新方向

AI技术在肝病临床试验方法学领域正处于探索发展阶段,其应用范围与深度仍然有限。目前,AI主要以辅助工具形式嵌入数据处理、患者筛选、风险分层和终点评价等环节,这类应用多聚焦于数据挖掘和决策支持,而真正以AI为核心干预对象或验证内容的前瞻性临床试验仍然较为有限。AI辅助型与AI验证型肝病临床试验在设计方法、评估指标和方法学要求上各具特点,有待进一步规范。总体来看,AI相关创新型试验设计和流程优化已初步展现临床应用前景,但高质量循证证据尚不充足,亟需更多大规模、多中心的真实世界研究以推动其规范化和临床转化。

未来,肝病临床试验中AI方法学的优化路径主要包括以下方面:建立规范化数据采集与安全共享机制,确保数据质量与合规性;推动联邦学习等隐私计算技术在多中心研究中的应用,提升模型泛化能力;加强独立多中心外部验证研究,系统评估模型稳定性与临床适用性;深化多学科协作,整合AI、肝病学、流行病学、统计学及伦理学等领域专业知识,共同推动AI方法学标准的制

表3 AI临床试验中涉及伦理与合规的相关规范

Table 3 Principal ethical and regulatory frameworks for AI in clinical trials

规范名称(发布时间)	适用范围	核心内容
WHO: Ethics and governance of artificial intelligence for health (2024)	健康领域临床研究与试验中AI工具的开发和应用	透明度、公平性、数据隐私、受试者自主权
FDA: Proposed Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SAMD) (2019)	AI软件在临床试验及医疗应用中的开发、验证和修改	风险管理、软件生命周期管理、模型修改的监管要求、确保可信度
EMA: Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle (2023)	药物研发及临床试验中AI的应用,包括设计、验证和监管	透明性、可解释性、合规要求、监管及验证指导
FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare (2023)	医疗AI开发与应用全过程,尤其适用于临床试验阶段的设计、验证和部署	6项核心原则(公平性、普适性、可追溯性、可用性、稳健性、可解释性)+28项最佳实践

表4 AI医学研究报告与方法学规范

Table 4 Reporting and methodological standards for AI-based medical studies

规范名称(发布时间)	适用范围	适用的临床研究类型
CONSORT-AI(2020)	AI相关临床试验结果报告规范	随机对照试验(AI作为干预或辅助工具)
SPIRIT-AI(2020)	AI相关临床试验方案撰写与报告规范	随机对照试验(AI作为干预或辅助工具)
MINIMAR(2020)	医学AI研究最低报告要求及其跨领域补充	各类早期AI临床研究的基本标准
DECIDE-AI(2022)	AI决策支持系统的早期临床研究报告规范	可行性研究
TRIPOD+AI(2024)	个体预后或诊断多变量预测模型透明报告规范——AI扩展	个体层面的多变量预测模型研究
CLAIM(2024更新)	医学影像AI研究透明度与可重复性规范	影像学AI研究
STARD-AI(制定中)	AI诊断准确性研究报告规范	AI诊断准确性评价研究
QUADAS-AI(制定中)	AI诊断准确性研究的偏倚风险评估工具	AI诊断研究的方法学质量评价
PROBAST+AI(制定中)	AI预测模型研究的偏倚风险与适用性评估	AI模型研究的方法学质量评价

注:数据来源于EQUATOR Network(<https://www.equator-network.org/>)。

定与实施。随着AI方法学的不断完善及配套规范体系的建立, AI技术将加速融入肝病临床试验, 助力该领域实现高质量、高效率 and 智能化发展。

利益冲突声明: 本文不存在任何利益冲突。

作者贡献声明: 孔媛媛负责设计论文框架, 起草与修改论文并最后定稿; 史立臣负责研究过程的实施, 起草与修改论文; 史立臣和冯睿时负责数据收集, 绘制表格; 尤红和贾继东负责指导论文修改。

参考文献:

- [1] KANG HJ, LEE JM, AHN C, et al. Low dose of contrast agent and low radiation liver computed tomography with deep-learning-based contrast boosting model in participants at high-risk for hepatocellular carcinoma: Prospective, randomized, double-blind study [J]. *Eur Radiol*, 2023, 33(5): 3660-3670. DOI: 10.1007/s00330-023-09520-4.
- [2] BAE JS, LEE JM, KIM SW, et al. Low-contrast-dose liver CT using low monoenergetic images with deep learning-based denoising for assessing hepatocellular carcinoma: A randomized controlled non-inferiority trial[J]. *Eur Radiol*, 2023, 33(6): 4344-4354. DOI: 10.1007/s00330-022-09298-x.
- [3] YOON JH, PARK JY, LEE SM, et al. Renal protection CT protocol using low-dose and low-concentration iodine contrast medium in at-risk patients of HCC and with chronic kidney disease: A randomized controlled non-inferiority trial[J]. *Cancer Imaging*, 2023, 23(1): 100. DOI: 10.1186/s40644-023-00616-0.
- [4] ODISIO BC, ALBUQUERQUE J, LIN YM, et al. Software-based versus visual assessment of the minimal ablative margin in patients with liver tumours undergoing percutaneous thermal ablation (COVER-ALL): A randomised phase 2 trial[J]. *Lancet Gastroenterol Hepatol*, 2025, 10(5): 442-451. DOI: 10.1016/S2468-1253(25)00024-X.
- [5] JEON SK, LEE JM, JOO I, et al. Two-dimensional convolutional neural network using quantitative US for noninvasive assessment of hepatic steatosis in NAFLD[J]. *Radiology*, 2023, 307(1): e221510. DOI: 10.1148/radiol.221510.
- [6] WU CH, YEN KC, WANG LY, et al. Automated whole-liver fat quantification with magnetic resonance imaging-derived proton density fat fraction map: A prospective study in Taiwan[J]. *Gut Liver*, 2025, 19(4): 617-626. DOI: 10.5009/gnl240408.
- [7] JOSHI S, SHAMANNA P, DHARMALINGAM M, et al. Digital twin-enabled personalized nutrition improves metabolic dysfunction-associated fatty liver disease in type 2 diabetes: Results of a 1-year randomized controlled study[J]. *Endocr Pract*, 2023, 29(12): 960-970. DOI: 10.1016/j.eprac.2023.08.016.
- [8] TIYARATTANACHAI T, APIPARAKOON T, CHAICHUEN O, et al. Artificial intelligence assists operators in real-time detection of focal liver lesions during ultrasound: A randomized controlled study[J]. *Eur J Radiol*, 2023, 165: 110932. DOI: 10.1016/j.ejrad.2023.110932.
- [9] LEE DH, LEE JM, LEE CH, et al. Image quality and diagnostic performance of low-dose liver CT with deep learning reconstruction versus standard-dose CT[J]. *Radiol Artif Intell*, 2024, 6(2): e230192. DOI: 10.1148/ryai.230192.
- [10] ZHANG YM, CUI J, WAN W, et al. Multimodal imaging under artificial intelligence algorithm for the diagnosis of liver cancer and its relationship with expressions of EZH2 and p57[J]. *Comput Intell Neurosci*, 2022, 2022: 4081654. DOI: 10.1155/2022/4081654.
- [11] KANG M, KIM MS. Managing postembolization syndrome through a machine learning-based clinical decision support system: A randomized controlled trial[J]. *Comput Inform Nurs*, 2024, 42(11): 817-828. DOI: 10.1097/CIN.0000000000001188.
- [12] LOOMBA R, NOUREDDIN M, KOWDLEY KV, et al. Combination therapies including cilofexor and firsocostat for bridging fibrosis and cirrhosis attributable to NASH[J]. *Hepatology*, 2021, 73(2): 625-643. DOI: 10.1002/hep.31622.
- [13] MCMICHAEL J, LIEBERMAN R, MCCAULEY J, et al. Computer-guided randomized concentration-controlled trials of tacrolimus in autoimmunity: Multiple sclerosis and primary biliary cirrhosis[J]. *Ther Drug Monit*, 1996, 18(4): 435-437. DOI: 10.1097/00007691-199608000-00021.
- [14] FENG SJ, WANG JH, WANG LH, et al. Current status and analysis of machine learning in hepatocellular carcinoma[J]. *J Clin Transl Hepatol*, 2023, 11(5): 1184-1191. DOI: 10.14218/JCTH.2022.000775.
- [15] CALDERARO J, ŽIGUTYTĖ L, TRUHN D, et al. Artificial intelligence in liver cancer: New tools for research and patient management[J]. *Nat Rev Gastroenterol Hepatol*, 2024, 21(8): 585-599. DOI: 10.1038/s41575-024-00919-y.
- [16] GHOSH S, ZHAO X, ALIM M, et al. Artificial intelligence applied to omics data in liver disease: Towards a personalised approach for diagnosis, prognosis and treatment[J]. *Gut*, 2025, 74(2): e331740. DOI: 10.1136/gutjnl-2023-331740.
- [17] CALDERARO J, SERAPHIN TP, LUEDDE T, et al. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma[J]. *J Hepatol*, 2022, 76(6): 1348-1361. DOI: 10.1016/j.jhep.2022.01.014.
- [18] MATTHAY EC, NEILL DB, TITUS AR, et al. Integrating artificial intelligence into causal research in epidemiology[J]. *Curr Epidemiol Rep*, 2025, 12(1): 6. DOI: 10.1007/s40471-025-00359-5.
- [19] CLUSMANN J, BALAGUER-MONTERO M, BASSEGODA O, et al. The barriers for uptake of artificial intelligence in hepatology and how to overcome them[J]. *J Hepatol*, 2025 DOI: 10.1016/j.jhep.2025.07.003. [Online ahead of print]

收稿日期: 2025-08-25; 录用日期: 2025-10-08

本文编辑: 刘晓红

引证本文: SHI LC, FENG RS, JIA JD, et al. Application of artificial intelligence in clinical trials of liver diseases: A methodological perspective[J]. *J Clin Hepatol*, 2025, 41(11): 2227-2234. 史立臣, 冯睿时, 贾继东, 等. 人工智能相关肝病临床试验的方法学考量[J]. *临床肝胆病杂志*, 2025, 41(11): 2227-2234.