

DOI:10.13288/j.11-2166/r.2023.21.002

基于中医临床诊疗数据的真实世界及人工智能研究思路与挑战

赵国桢¹, 郭诗琪², 庞华鑫³, 高子恒⁴, 李博⁴, 蔡兆伦⁵,
闫世艳², 韩东燃², 刘一星², 胡晶⁴, 刘清泉⁴✉

1. 中国中医科学院中医临床基础医学研究所, 北京市东城区东直门内南小街16号, 100700; 2. 北京中医药大学; 3. 北京交通大学计算机与信息技术学院; 4. 首都医科大学附属北京中医医院/北京市中医药研究所北京中医药循证医学中心; 5. 四川大学华西医院

[摘要] 随着真实世界研究方法学的进步和人工智能技术的成熟, 以中医诊疗数据为基础, 开展“定量”研究指导临床实践的方法正在逐步形成, 但在研究总体设计、临床转化等方面仍需进一步完善。提出基于中医临床诊疗数据的真实世界及人工智能研究的总体设计思路及应用方法, 包括5个步骤: 建设高质量大样本的具有中医特色的研究型数据库、核心处方的挖掘及分类、队列研究评价核心处方疗效、病例对照研究明确优势人群、建立预测模型实现精准医疗。此外, 研究人员还需要建立标准化的中医变量采集和数据处理体系, 优化混杂因素的确定及测量方法, 进一步完善及推广方法学, 加强跨学科人才的培养。真实世界研究与人工智能研究的研究基础相似, 临床应用互补, 未来将是二者的融合。

[关键词] 人工智能; 真实世界研究; 大数据; 临床研究方法

随着真实世界研究 (real-world study, RWS) 的提出和实践, 一系列使用中医临床诊疗数据开展研究并指导临床实践的方法逐步形成, 推动了中医临床研究方法学的发展^[1-2]。同时, ChatGPT 和 GPT-4 等人工智能 (artificial intelligence, AI) 技术的问世, 让更多学者关注到基于大数据的人工智能技术对医疗行业的影响^[3]。中医学理论体系与西医学理论体系差异巨大, 因而需要建立基于中医学特点的研究模式, 为中医学的传承和发展提供动力。中医学是通过大量临床实践传承和发展的学科, 医家学者不断总结名老中医临床诊疗经验, 逐渐形成自身的学术观点、核心处方和用药特点, 并在实践中进一步完善和创新^[4], 这种基于“定性”的传承和发展模式是目前的主流模式。随着研究方法的进步、人工智能技术的成熟以及多种测量工具的出现, 以采集中医临床诊疗数据, 建立研究型数

据库, 开展真实世界和人工智能研究, 最终指导临床实践的“定量”研究方法将逐步形成, 但相关方法学仍不完善, 主要体现在缺少研究总体设计和临床转化方法等方面。针对这一问题, 我们提出基于中医临床诊疗数据的真实世界及人工智能研究的总体设计思路及应用方法, 以期进一步推动中医临床研究方法学的发展和研究成果的临床转化。

1 中医真实世界及人工智能研究概述

1.1 中医 RWS 的现状存在问题

RWS 包含多种研究类型, 相较于解释性随机对照试验 (explanatory randomized control study, eRCT) 具有研究目的多样、纳入及排除标准宽泛、外推性强等优势, 且可使用临床诊疗数据开展研究, 与中医学传承发展模式高度契合^[5]。近年来中医 RWS 发表数量逐年增加, 尤其以回顾性队列研究为主, 并为重大突发疾病的临床用药快速提供了可靠证据^[6]。然而, 真实世界数据来源多样、人群异质性高, 使 RWS 往往存在更多的偏倚, 因而需要更严谨的研究设计和更复杂的统计分析方法^[7]。

基金项目: 中国中医科学院科技创新工程项目资助(CI2021A00702-3); 国家中医药多学科交叉创新团队项目(ZYYCXTD-D-202201)

✉ 通讯作者: liuqingquan2003@126.com

此外,在中医临床诊疗中,医生会根据患者体质、中医证型和病情变化,灵活使用草药处方、针刺穴位等中医治疗措施,使队列研究中的暴露因素不再单一或固定,且更易受时依混杂的影响,故无法完全照搬常规的西医研究方法^[8]。这需要基于中医特色,探索出能够符合中医特点、为临床实践真正提供指导意义的研究模式。

1.2 中医人工智能研究的现状及存在问题

人工智能技术即研究者编写计算机程序指令,使计算机对具体问题做出响应,提供具有“智慧”的决策方案。人工智能技术需要对大量数据进行学习分析,常用的模型包括线性回归、Lasso 回归、K 最近邻、贝叶斯网络、决策树、随机森林、梯度提升树等^[9-10]。此外,以仿人类神经运行模式的人工神经网络算法为代表的深度学习技术正快速发展。如在新型冠状病毒感染疫情中,深度学习技术辅助医生对肺部 CT 进行快速分析和诊断,提高了临床工作效率和诊疗水平^[11]。

中医学领域也开展了大量人工智能研究,涵盖了“诊”与“治”,如智能中医四诊仪、智能中医开方系统等^[12-13]。但值得关注的是,人工智能技术的本质是统计学和计算机语言,本身不具有逻辑思维,而目前公开的中医人工智能技术大多将变量通过某种算法简单链接,加上神经网络的应用,使现有算法严重缺乏可解释性,存在脱离中医理论及临床实践的风险^[14]。为推动中医领域人工智能研究,应在中医理论指导下构建符合中医诊疗规律的人工智能模型及算法策略,以尽可能增强可解释性。此外,还应对人工智能技术的可靠性和有效性进行评价,为其临床推广应用提供证据。

1.3 RWS 与人工智能研究的关系

RWS 是基于真实世界数据,通过统计分析,形成真实世界证据,主要用于药物的审评审批,也可临床诊疗提供证据^[15]。而人工智能研究则是数据与模型相辅相成,基于数据的特点构建合适的临床诊疗模型架构。模型通过训练和迭代优化,高效且相对准确地生成诊疗方案,可直接辅助临床决策^[16]。

在理论基础层面,RWS 和人工智能研究都是以数理方法为指导。在数据基础层面,RWS 与人工智能研究均基于医疗数据开展。但 RWS 更强调数据来源,要求数据在真实世界环境下收集;而人工智能研究对数据来源则无明确要求,也可以使用随机对照试验数据建立模型^[17]。在临床应用层面,

二者均能辅助临床医生开展诊疗活动,但 RWS 更倾向于提供潜在获益-风险的临床证据,其可靠度较高,使用效率较低;而人工智能研究则倾向于建立临床辅助决策系统,提高诊疗效率,但其可靠度仍需进一步评价。

因此,RWS 与人工智能研究的研究基础相似,临床应用互补,未来研究趋势将是二者的融合。基于多维度的真实世界证据,恰当地引入人工智能模型,智能辅助临床决策;再对基于人工智能技术的临床辅助决策系统开展 RWS,对其可靠度进行评价。

2 中医真实世界及人工智能研究的总体设计思路及应用方法

2.1 建设高质量大样本的研究型数据库

数据库需基于研究目的、疾病特点进行针对性设计,如中医数据库通常需要包含患者的舌脉数据。数据库的高质量体现在数据的完整性、准确性、透明性,并在数据治理的各个环节开展质量控制^[15]。完整性是指数据的缺失程度。数据缺失难以避免,但可通过数据溯源、研究人员培训、患者关怀等方式,按时定期随访,尽可能减少缺失;对于无法避免的缺失可通过统计方法填补。准确性是指数据与其描述的客观特征是否一致。每一数据都应有可靠的来源,能够客观准确地记录患者的真实情况,不应使用随意复制的数据,或未经测量随意填写的数据。透明性是指数据治理方案和过程透明。应保留原始数据,数据采集、核查和清洗过程记录,以及衍生变量的计算方法等。

数据库的样本量应满足多种 RWS 统计分析及人工智能算法的需要。不同类型的 RWS 都有各自的样本量计算公式,同时还需考虑限制、匹配等方法引起的样本量减少,以及多因素分析的样本量要求^[18]。深度学习算法同样存在样本量要求,如神经网络算法的样本量不少于输入层变量数量的 200 倍,且样本量越大,模型越稳定^[19]。

2.2 挖掘及分类核心处方

真实世界环境中,不同中药处方的药物组成存在差异,为后续研究带来困难,因此需对核心处方进行挖掘和分类。核心处方的挖掘推荐采用混合研究方法。首先,通过无监督学习算法探索处方用药规律,初步挖掘核心处方。随后,根据挖掘结果,结合专家访谈、问卷调查、临证经验梳理等定性研究,明确数据库中的多个核心处方。通常情况下,同一疾病的核心处方可能多达数十个。最后,通过

相似度匹配算法对真实世界核心处方进行分类,以“类方”的形式开展进一步研究^[20]。

2.3 队列研究评价核心处方疗效

队列研究是 RWS 中较常见的研究类型,主要用于评价临床疗效和检验病因假设。其根据研究对象是否暴露于某个待研究的危险因素分组,通过对比各组结局的发生,推断暴露因素(治疗措施)与疾病(结局)之间的因果关系^[21]。作为随机对照试验(RCT)的补充,队列研究契合中医学“先临床后研究”的特点,可以使用治理后的临床诊疗数据,快速对多种不同治疗措施进行疗效评价。例如,在新型冠状病毒感染疫情早期,基于临床诊疗数据发表了多篇高影响力的回顾性队列研究,为抗疫用药及时提供证据^[6,22]。

需注意的是,混杂因素和偏倚始终是队列研究不可忽视的局限性^[5]。目前,对于混杂因素常用的统计处理方法包括分层、倾向性评分和多因素回归等^[23]。虽然在研究设计、数据管理、统计分析的各个阶段都会对偏倚和混杂进行控制,但偏倚和混杂始终无法避免。我们认为,通常情况下(排除罕见病等特殊情况)队列研究更适用于多种中医药治疗措施的初步评价,从大量的核心处方、中成药中快速筛选出可能有效的治疗措施。在明确各治疗措施的优势人群特征后,仍需要通过 RCT 进一步评价确切疗效。

2.4 病例对照研究明确优势人群

2012 年,刘建平教授与 Vinjar F 教授提出了中医最佳—最差病例的研究思路,并逐渐发展成为优势人群特征分析^[24]。该方法本质上属于病例对照研究,根据接受指定治疗措施后是否发生某个结局,将患者分为优势组和劣势组,通过对比明确使用该治疗措施有效的人群特征。有学者通过病例对照研究,发现结直肠癌患者的中医优势人群特征为血瘀质体质的患者,劣势人群特征为阳虚质体质或肝脾不和证的患者^[25]。优势人群特征分析将成为指导临床实践及精准医疗的重要研究方法之一。基于优势人群特征分析结果,可为 RCT 提供更加精准的研究对象筛选标准,弥补以往中医 RCT 中因辨证不准确可能引起的低估药物真实疗效的问题。

2.5 建立预测模型实现精准医疗

中医临床治疗方案选择多样,医生可选择草药处方、中成药或中医技术中的一种或多种,可单独中医治疗或与西医联合治疗。治疗方案的选择在临床实践中高度依赖知识水平和临床经验,通常缺少

定量研究的数据支持。

随着精准医学发展,基于临床诊疗数据预测治疗措施疗效的预测模型被提出,如反事实预测模型^[26-27]。该模型是预后模型的一种,其要点是部分预测因子是未发生的事件,即反事实事件。应用场景举例:某患者若接收了某种治疗,未来出现转重的概率是多少?这类预测模型可基于中医理论构建,通过预测及对比患者在不同用药方案后的预后情况,推荐最适宜的用药方案,为临床精准用药提供参考。

此外,随着人工智能技术的发展,预测能力更强、可处理复杂网络及多维数据的模型及算法逐渐被创造及应用^[28-29]。但无论哪种模型或算法被应用到中医临床诊疗中,都应遵循中医理论框架,符合中医临床思维。例如,将患者主观感受、舌脉等中医诊疗信息纳入模型,实现结果的可解释性。此外,在建立模型后,除了使用验证集对模型开展评价,还需进一步开展临床研究对模型进行有效性、安全性及可靠性评价。

3 开展中医真实世界及人工智能研究的挑战及对策

3.1 中医变量的采集及数据处理

中医变量是在中医药理论指导下,可以反映中医学症状、体征及内在变化规律等特征的变量,如中医体质、舌脉、面色、主观感受等^[30]。其中,部分中医变量既可以影响中医医生治疗用药的选择,又会影响患者的转归结局,且不是因果链上的中间变量,这些中医变量在 RWS 中属于混杂因素,在人工智能研究中属于预测因子^[31]。多数已发表的中医 RWS 仅选择患者生命体征、西药使用情况和实验室检查等西医变量作为混杂因素,通常缺少中医变量的应用;部分研究虽然使用了中医变量,但不同研究间对中医变量的采集和数据处理方法存在差异^[6,22]。

有研究者提出的中医主观表型组学体系,可通过小程序采集患者的 432 个生理特征、320 个心理特征和 267 个生活习惯特征数据,初步实现对人体主观感受的采集^[32]。有研究使用智能化设备采集患者舌诊、面诊图像,并通过红绿蓝(RGB)、色调-饱和度-亮度(HSV)、亮度-a-b 通道(Lab)等不同颜色模型进行数据处理和图像显示,实现了舌像、面像信息的量化测量^[33-37]。对于脉诊数据,有研究通过压力感受器绘制脉图,并基于主波幅度(h1)、主波上 1/3 处的宽度(w)、脉搏周期(t)、

脉搏周期面积 (s) 等变量处理数据, 初步实现脉诊信息的量化^[38]。目前中医变量的采集和数据处理方法仍处于探索阶段, 尚缺少公认的标准, 未来需基于临床实践验证其实用性及可靠性并予以完善, 逐步建立中医变量的标准化采集和数据处理体系^[12]。

3.2 混杂因素的确定及测量

中医 RWS 往往需要先建立因果推断模型证明其确切疗效, 而混杂因素对疗效的因果推断有着较大的干扰, 是观察性疗效比较研究中不可忽视的重要变量。倾向性评分、疾病风险评分法等统计学方法常被用于控制混杂, 然而确定并选择混杂因素同样重要^[39]。在观察不同中医治疗措施治疗相同疾病的疗效比较研究中, 混杂因素理论上应该是一致的, 但不同研究选择的混杂因素通常存在较大差异。因此, 我们可借鉴核心结局指标集的研究思路, 根据不同疾病特点, 通过文献检索、专家访谈、专家共识等方法, 明确影响各个疾病疗效的主要混杂因素, 形成“核心混杂因素集”, 并使用有向无环图 (directed acyeling graph, DAG) 客观反映各混杂因素的关系, 以便于研究人员在相关疾病的队列研究中参考和使用^[31]。

此外, 还有不少已知但尚无测量方法的混杂因素, 例如患者管理策略、护理水平、医生操作手法、医生是否对患者进行心理疏导等, 以及部分未知混杂因素, 均会影响最终的疗效。已有学者针对这一问题开展探索, 如使用 E 值评估未知混杂大小^[40-41]。未来可进一步设计针对性的测量工具, 形成标准化的测量方法, 并在 RWS 中探索和应用。

3.3 方法学的完善及推广

目前, 中医领域真实世界及人工智能研究的方法学仍不够完善, 且推广力度不足。例如, 在部分发表的队列研究中, 未考虑混杂因素对结局的影响, 或未使用正确的统计分析方法对混杂因素进行控制和处理。建议相关部门加强对中医真实世界及人工智能方法学研究的资助力度, 实行鼓励中文方法学论文发表的政策, 增加专业的培训机构和公开课的数量, 以及鼓励开发临床研究辅助软件, 降低方法学学习成本。

3.4 跨学科人才的培养

真实世界及人工智能研究涉及的学科较多, 包括中医类临床医学、循证医学、流行病学、信息学及统计学等。为提高研究质量, 推动成果临床转化, 需要一批既熟悉中医临床思维、了解中医诊疗

现状, 又掌握多种研究方法技术的跨学科人才。建议在开展相关研究的同时, 注重对青年人才的培养, 建立完善的跨学科人才培养机制, 设立跨学科人才培养基金, 为未来的研究工作做好人才储备。

4 结语

结合中医学科“先临床后研究”的特点, 随着研究方法的完善和技术的进步, 基于中医临床诊疗数据开展真实世界及人工智能研究已成为未来中医学发展的重要方向之一。本文所提出的总体研究思路可概括为: 第一步, 使用客观且标准的方法采集数据, 经过严格的数据治理, 建立中医特色的研究型数据库; 第二步, 通过混合方法研究挖掘核心处方, 使用相似度匹配算法对真实世界中药处方进行分类; 第三步, 采用队列研究设计, 在识别及控制混杂因素的基础上, 对各个核心处方或中成药的疗效进行评价; 第四步, 采用病例对照设计, 对潜在有效的核心处方和中成药开展优势人群特征分析, 为精准用药提供证据; 第五步, 构建预测模型, 为临床精准用药提供参考。此外, 为高质量地开展相关研究, 仍需要技术上的创新, 以实现标准化的中医变量采集及数据处理; 需要方法学的完善和推广, 明确混杂因素及探索混杂的测量方法; 还需要重视方法学的应用及跨学科人才的培养。这一研究模式有望推动研究成果临床转化, 促进中医精准医疗的发展, 真正实现“解码中医”。

参考文献

- [1] SUN X, TAN J, TANG L, et al. Real world evidence: experience and lessons from China[J]. BMJ, 2018, 360: j5262. doi: 10.1136/bmj.j5262.
- [2] 刘保延. 真实世界的中医临床科研范式[J]. 中医杂志, 2013, 54(6): 451-455.
- [3] LEE P, BUBECK S, PETRO J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine[J]. N Engl J Med, 2023, 388(13): 1233-1239.
- [4] 谷晓红, 于河, 卜德超, 等. 应用混合方法构建名老中医道术传承研究与推广范式[J]. 中医杂志, 2022, 63(15): 1406-1411.
- [5] 吴阶平医学基金会中国胸部肿瘤研究写作组. 真实世界研究指南(2018年版) [EB/OL]. (2018-09-20) [2023-04-05]. <http://www.xayxfw.com/uploadfiles/2018/09/201809071442494249.pdf>.
- [6] TIAN J, YAN S, WANG H, et al. Hanshiyi Formula, a medicine for Sars-CoV2 infection in China, reduced the proportion of mild and moderate COVID-19 patients turn-

- ing to severe status: a cohort study[J]. *Pharmacol Res*, 2020, 161: 105127. doi: 10.1016/j.phrs.2020.105127.
- [7] MOONS KG, ROYSTON P, VERGOUWE Y, et al. Prognosis and prognostic research: what, why, and how? [J]. *BMJ*, 2009, 338: b375. doi: 10.1136/bmj.b375.
- [8] 梁洁, 和思敏, 陈淑婷, 等. 纵向研究中控制时依混杂的 G 方法 [J]. *中华流行病学杂志*, 2021, 42(10): 1871-1875.
- [9] 田之魁, 张鑫海, 孙璇, 等. 基于舌象的糖尿病足风险预测模型的构建 [J]. *中医杂志*, 2022, 63(19): 1840-1846.
- [10] 高群, 张华军, 张立晶, 等. 基于十二井穴生物电阻抗值构建冠心病患者抗血栓治疗出血风险预测模型 [J]. *中医杂志*, 2023, 64(7): 710-715.
- [11] HUANG L, HAN R, AI T, et al. Serial quantitative chest CT assessment of COVID-19: a deep learning approach [J]. *Radiol Cardiothorac Imaging*, 2020, 2(2): e200075.
- [12] 赵国桢, 郭诗琪, 庞华鑫, 等. 人工智能技术在辅助中医诊疗及诊疗标准化中的应用 [J]. *中医杂志*, 2022, 63(24): 2306-2310.
- [13] 王阶, 李谦一, 张振鹏, 等. 基于循环神经网络的智能中医开方模型训练方法及装置: CN112466436A [P]. 2021-03-09.
- [14] 张姝艳, 皮婷婷. 医疗领域中人工智能应用的可解释性困境与治理 [J]. *医学与哲学*, 2023, 44(3): 25-29, 35.
- [15] 国家药品监督管理局药品审评中心. 国家药监局药审中心关于发布《用于产生真实世界证据的真实世界数据指导原则(试行)》的通告 [EB/OL]. (2020-04-15) [2023-04-15]. <https://www.cde.org.cn/zdyz/domesticinfopage?zdyzIdCODE=7d2e46cea0e459358257760383526e9d>.
- [16] 田贵华, 商洪才. 智能中医学概论 [M]. 北京: 人民卫生出版社, 2021: 132-133.
- [17] PAJOUHESHNI R, GROENWOLD RHH, PEELEN LM, et al. When and how to use data from randomised trials to develop or validate prognostic models [J]. *BMJ*, 2019, 365: l2154. doi: 10.1136/bmj.l2154.
- [18] 王树月, 陈明珠, 张兴会, 等. 药物流行病学研究混杂因素的控制 [J]. *药物流行病学杂志*, 2021, 30(8): 503-507.
- [19] RILEY RD, ENSOR J, SNELL KIE, et al. Calculating the sample size required for developing a clinical prediction model [J]. *BMJ*, 2020, 368: m441. doi: 10.1136/bmj.m441.
- [20] 赵国桢, 卢海天, 闫世艳, 等. 基于相似度匹配算法对真实世界草药处方分类的探索及实例解读 [J]. *中国中药杂志*, 2023, 48(4): 1132-1136.
- [21] 范红敏. 队列研究在医学科研中的理论、设计和实践 [J]. *北京医学*, 2020, 42(8): 756-761.
- [22] ZHANG Y, LIU Y, LI M, et al. Timely treatment and higher compliance to traditional Chinese medicine: new influencing factors for reducing severe COVID-19 based on retrospective cohorts in 2020 and 2021 [J]. *Pharmacol Res*, 2022, 178: 106174. doi: 10.1016/j.phrs.2022.106174.
- [23] JAGER KJ, ZOCCALI C, MACLEOD A, et al. Confounding: what it is and how to deal with it [J]. *Kidney Int*, 2008, 73(3): 256-260.
- [24] 杨宇飞. 晚期消化道恶性肿瘤中医优势人群特征及诊疗路径研究 [J]. *北京中医药大学学报*, 2022, 45(5): 433-437.
- [25] 孙凌云, 张达, 郝洁, 等. 晚期结直肠癌患者中医证型、中医体质与生存结局相关性分析 [J]. *中国中西医结合杂志*, 2017, 37(9): 1059-1062.
- [26] VAN GELOVEN N, SWANSON SA, RAMSPEK CL, et al. Prediction meets causal inference: the role of treatment in clinical prediction models [J]. *Eur J Epidemiol*, 2020, 35(7): 619-630.
- [27] 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.
- [28] YUAN L, YANG L, ZHANG S, et al. Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: a prospective multicentre clinical cohort study [J]. *E Clinical Medicine*, 2023, 57: 101834. doi: 10.1016/j.eclinm.2023.101834.
- [29] WANG Y, XU J, ZHANG J, et al. SIAP: an intelligent algorithm for multiple prescription pattern recognition based on weighted similarity distances [J]. *BMC Med Inform Decis Mak*, 2023, 23(1): 79.
- [30] 吴涛涛, 王琦, 李岩, 等. 基于体病相关探讨临床预测模型的构建 [J]. *中医杂志*, 2022, 63(16): 1507-1510, 1531.
- [31] 刘颖, 刘旭, 汪颖, 等. 《因果推断研究中混杂因素控制与结果报告指南》解读 [J]. *中华危重病急救医学*, 2021, 33(1): 113-116.
- [32] 落楠. 探寻中医药疗效评价新范式 [N]. *中国医药报*, 2022-12-14(004).
- [33] FENG L, HUANG ZH, ZHONG YM, et al. Research and application of tongue and face diagnosis based on deep learning [J]. *Digit Health*, 2022, 8: 20552076221124436. doi: 10.1177/20552076221124436.

- [34]张婷婷,王亚强,温川飙,等. 论智能中医辨证可解释研究的挑战和研究思路[J]. 中医杂志, 2022, 63(6): 506-510.
- [35]周明瀚,刘旺华,李花,等. 原发性高血压阴虚阳亢证舌象客观化研究[J]. 中华中医药杂志, 2022, 37(6): 3401-3404.
- [36]陈清光,蔡梦洁,袁培琼,等. 不同糖代谢人群的舌象客观化特征研究[J]. 辽宁中医杂志, 2021, 48(3): 88-91.
- [37]关茜,杨帅,郭元成,等. 阴虚、阳虚体质人群的舌诊图像特征分析[J]. 中华中医药杂志, 2021, 36(8): 4860-4862.
- [38]崔骥,费兆馥,张志枫,等. 平、滑、弦脉的脉图分类参数范围研究[J]. 上海中医药杂志, 2016, 50(12): 15-17.
- [39]黄丽红,魏永越,陈峰. 如何控制观察性疗效比较研究中的混杂因素:(一)已测量混杂因素的统计学分析方法[J]. 中华流行病学杂志, 2019, 40(10): 1304-1309.
- [40]洪翔,尹洁晨,王蓓. “探针变量”:观察性研究中混杂因素识别的工具[J]. 中华流行病学杂志, 2021, 42(4): 735-739.
- [41]VANDER-WEELE TJ, DING P. Sensitivity analysis in observational research: introducing the E-Value [J]. Ann Intern Med, 2017, 167(4): 268-274.

Research Ideas and Challenge of Real World Study and Artificial Intelligence Based On Clinical Diagnosis and Treatment Data of Traditional Chinese Medicine

ZHAO Guozhen¹, GUO Shiqi², PANG Huaxin³, GAO Ziheng⁴, LI Bo⁴, CAI Zhaolun⁵, YAN Shiyan², HAN Dongran², LIU Yixing², HU Jing⁴, LIU Qingquan⁴

1. Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing, 100700; 2. Beijing University of Chinese Medicine; 3. School of Computer and Information Technology, Beijing Jiaotong University; 4. Beijing Evidence-based Chinese Medicine Center, Beijing Traditional Chinese Medicine Hospital, Capital Medical University/ Beijing Institute of Chinese Medicine; 5. West China Hospital, Sichuan University

ABSTRACT With the continuous progress of research methodology in the real world and the growing maturity of artificial intelligence technology, a method for conducting “quantitative” research to guide clinical practice based on traditional Chinese medicine (TCM) diagnosis and treatment data was gradually developed. However, there is still a need for further improvements in the overall design of studies and the transformation of findings into clinical practice. Based on this, we put forward a comprehensive overall design concept and application approach for real-world study and artificial intelligence research based on clinical diagnosis and treatment data of TCM. This approach consists of five steps: Constructing a research-based database with a large sample size and high data quality; Mining and classification of core prescriptions; Conducting cohort studies to evaluate the effectiveness of core prescriptions; Utilizing case-control studies to clarify the dominant population; Establishing predictive models to achieve precision medicine. Additionally, it is imperative for researchers to establish a standardized system for collecting TCM variables and processing data, optimize the determination and measurement methods of confounding factors, further improve and promote methodologies, and strengthen the training of interdisciplinary talents. By following this research method, we anticipate that the clinical translation of research findings will be facilitated, leading to advancements in TCM precision medicine. Real-world study and artificial intelligence research share similar research foundations, and clinical applications complement each other. In the future, the two will merge together.

Keywords artificial intelligence; real world study; big data; method of clinical research

(收稿日期: 2023-05-06; 修回日期: 2023-08-28)

[编辑: 柴倩云]