

中医优势病种专栏

[专栏导读] 社区获得性肺炎作为临床最常见的呼吸系统感染性疾病, 发病率居高不下, 严重威胁人民群众生命健康, 且面临着病原谱变迁、细菌耐药、迁延难愈等临床困境。中医药在社区获得性肺炎的防治中历经数千年临床实践, 形成了系统的理论体系与丰富的诊疗经验, 以辨证论治为核心的个体化诊疗模式展现出独特的临床优势。随着真实世界研究方法的持续深化、大数据技术的广泛应用以及人工智能技术的飞速发展, 尤其是多模态大模型在医疗领域的深度落地, 中医药治疗社区获得性肺炎的研究正逐步从传统经验总结转向数据驱动的循证研究, 由单一的临床疗效观察迈向涵盖平台构建、数据解析、疗效验证、卫生经济学评价的多维度工程, 为论证中医药治疗本病的特色与优势提供了新的研究路径。

本专栏聚焦社区获得性肺炎的中医临床诊疗与研究前沿, 依托江苏省中医优势病种诊疗大数据平台, 遴选多篇具有代表性的研究论文, 内容涵盖多模态大模型赋能的平台建设与专病队列构建、参考中医优势病种诊疗方案辨证施治的真实世界研究及卫生经济学评价等方向, 从研究路径的构建, 诊疗方案的验证, 再到方剂的解析, 层层递进、互为支撑, 系统呈现了人工智能赋能、多中心数据驱动下中医药治疗社区获得性肺炎的潜力。本专栏的推出, 旨在系统呈现中医药治疗社区获得性肺炎的研究进展与深层思考, 推动中医优势病种建设, 完善中医诊疗规范, 助力中医药循证证据体系的构建与现代化发展。

DOI:10.13288/j.11-2166/r.2026.09.008

基于多模态大模型中医大数据平台的多中心 社区获得性肺炎真实世界队列的构建与应用

王志超¹, 周贤梅¹, 冯凡超¹, 王梦琪², 王鑫³, 康彬⁴, 俞晓帆⁵, 王晓骁¹,
肖蕾¹, 李娟¹, 张志超^{1,6}, 马叶^{1,6}, 季叶清^{1,6}, 童心⁷, 吴琢玥⁷, 刘佳²✉

1. 南京中医药大学附属医院, 江苏省南京市秦淮区汉中路155号, 210029; 2. 中国中医科学院中医临床基础医学研究所; 3. 中国中医科学院中医药数据中心; 4. 南京邮电大学物联网学院; 5. 南京邮电大学空天地海通信技术一体化研究院; 6. 南京中医药大学第一临床医学院; 7. 南京医科大学公共卫生学院

[摘要] 介绍依托江苏省中医优势病种诊疗大数据平台进行社区获得性肺炎(CAP)真实世界队列研究模式, 包括: 1) 数据清洗: 统一诊断、症状、治疗及影像信息, 智能化提取非结构化信息, 清洗并构建标准化数据库; 2) 队列构建: 按CAP诊断标准筛选全省CAP患者, 构建高质量专病队列; 3) 方案设计: 结合中医药研究特点及CAP病种特点, 确定合理的纳排标准、样本量估算方法、干预措施、结局指标、经济学评价指标等, 为CAP的中医真实世界研究提供参考。

[关键词] 社区获得性肺炎; 真实世界研究; 大数据平台; 人工智能

基金项目: 江苏省中医药科技发展计划(ZD202210, MS2021022);
中华中医药学会青年人才托举工程(2025-QNRC2-B03); 江苏省
中医院优秀青年博士培养计划(2023QB0103, 2023QB0108)

✉ 通讯作者: marie_liujia@sina.cn

近年来信息技术的飞速发展, 数字化、大数据和人工智能(artificial intelligence, AI)在医疗领域的运用^[1], 让原本难以处理的海量中医临床资料

逐步变成了可利用的真实世界数据,进一步推动了真实世界研究(real-world study, RWS)在中医药领域的应用^[2],为中医药的传承、创新和发展提供了新的机遇。与传统随机对照试验相比,RWS能够更好地反映中医药在真实临床环境中的有效性、安全性和经济性,具有贴近临床实际、疗效评价客观全面、安全监测长期稳定、辅助制定卫生决策等优势^[3]。为提高RWS的研究质量,国内外都在不断尝试建设真实世界数据平台,但目前平台建设多存在数据质量参差不齐、数据孤岛现象严重、数据转化效率低下、交互功能设计不足等问题^[4]。

江苏省中医药管理局于2023年启动建设江苏省中医优势病种诊疗大数据平台(以下简称平台),平台依托江苏省全民健康信息平台数据,融合了自然语言处理(NLP)、计算机视觉(CV)、大语言模型(LLM)等先进AI技术,形成医疗垂域的多模态大模型。平台核心数据架构由采集层、处理层、存储层、分析层及展示层共五个层级构成,各层级功能明确且协同联动,并全程由数据安全机制监测。平台首批优选了四个病种试点进行多中心RWS,社区获得性肺炎(community-acquired pneumonia, CAP)为其中之一。中医药可以结合证型及个体差异对CAP患者进行辨证施治,具有有效性、安全性、经济性、社会性、普遍性、地域性等特点^[5]。基于此,我们依托平台构建CAP真实世界大数据队列,为未来开展多中心、大样本的RWS提供数据支撑。

1 平台简介

1.1 平台基本构架

平台依托江苏省全民健康信息平台数据,融合了NLP、CV等先进AI技术,也包括近几年热门的LLM,形成医疗垂域的多模态大模型,分别构建中医电子病历、中医药基础资源、中医互联网医疗、中医药辅助决策四大数据库,汇聚了全省各级各类中医医疗机构的基本信息、科室情况、医疗人员、中医制剂、电子病历、术语字典等信息。平台能够支撑海量的多源异构数据的高效存储和处理,具备可用性、扩展性、实时性、数据自动标准化以及安全性。系统架构中嵌入了严格的权限管理、加密措施和安全审核机制。

1.2 数据类型及来源

患者临床数据来源主要包括门诊及住院病历、病案首页、医嘱、检验系统、检查系统、影像系

统、护理记录及随访系统等,通过AI技术进行智能化处理,形成丰富且标准化的数据类型,如流行病学、诊断、治疗、临床症状、问卷量表、检查指标、影像、随访、费用等信息。平台通过数据加密与访问控制机制保障数据安全的防线,采用数据去标识化与匿名化处理强化隐私保护,采用异常检测与安全响应体系实时监控处置风险,严格执行合规性与隐私保护标准,确保患者的数据安全。

2 基于平台的CAP真实世界队列的构建

2.1 CAP数据标准化

为确保数据的一致性、可比性及共享性,平台采用国际标准、国家标准与本地化标准相结合的办法,利用AI技术实现CAP中医病名、证候、证型、方剂、饮片等数据的统一编码与标准化处理。

2.1.1 诊断信息 采用《疾病和有关健康问题的国际统计分类:第十次修订本(ICD-10)》^[6]对西医诊断进行标准化,同时采用《中医病证分类与代码》^[7]对CAP中医诊断(病名、证型)进行标准化,对于没有采用上述标准的医院,可行信息化改造,或利用平台具备的机器学习与深度学习技术,通过知识图谱语义推理、图神经网络推理等算法,训练诊断相关的数据治理模型、数据格式化模型,将其他标准进行统一。如在临床关于邪袭肺卫证有多个同义或近义表达,经过训练后模型可自动将邪犯肺卫证、邪侵肺卫证、邪郁肺卫证、风邪犯肺证等归类为邪袭肺卫证。

2.1.2 症状信息 主要采用国际标准“系统化医学术语集——临床术语”(SNOMED CT)^[8]对病历系统进行半结构化处理,对于无法进行半结构化改造的医院,则在临床应用由专家制定与CAP研究对应的模板病历^[9]。病历数据汇集至平台后,利用Llama 3.3、DeepSeek R1为内核的LLM技术对文本进行多轮训练、微调和优化后,实现对文本超90%准确率的后结构化处理,覆盖病历中的发病时间、临床症状及严重程度、合并疾病、手术史、吸烟史、饮酒史、家族遗传史、预防接种史、婚育史等项目,形成标准化数据集。在批量处理万级数据时,支持分布式计算和多模型管理,减少计算负担,保持较高精度。

2.1.3 治疗信息 主要根据《中华人民共和国药典一部(2025年版)》^[10]、《国家基本药物目录(2012年版)》^[11]、《江苏省中药饮片炮制规范(2020年版 第一册)》^[12],利用NLP技术对医嘱信

息进行结构化处理。NLP 技术包括但不限于隐马尔可夫模型、条件随机场、维特比 (Viterbi) 算法、神经网络等, 能够基于文本知识标注, 针对专病角色、分词、句法、词性、上下文关系、语义等进行开发、建模, 从医嘱信息中提取关键治疗信息, 精准识别西药、中成药、中药饮片、外治疗法等治疗信息, 并对其进行归一和分类。由于各家医院中药饮片名称复杂多样, 单靠 NLP 技术无法完成数据标准化, 因此本项目对全省 2024 年度的 38 761 条中药饮片数据进行人工筛选、清洗、标记, 并组织专家进行了多轮论证, 形成了《江苏省中药饮片名称映射标准 (草案)》, 再通过 NLP 技术建模以达到对中药饮片信息标准化的目的。

2.1.4 影像信息 影像信息作为客观数据, 是很多病种临床研究的重要数据来源, 如 X 光片、CT、磁共振成像、超声、内镜等影像信息数据。在临床研究中如何实现多中心影像信息数据标准化, 直接影响研究结果的准确性。如 CAP 患者胸部 CT 扫描的核心诊断性检查, 也可以用作评估疗效。本研究利用 CV 技术可以实现对多中心胸部 CT 信息的标准化处理, 使二维、三维信息降维为可分析的数字型信息。为实现肺炎病灶区域分割和面积计算两大任务, 平台接入了一种以 U-Net 为主体框架、结合循环注意力反馈模块 (CAFM) 和博弈交互筛选模块 (GISM) 的新技术, 进一步提升了对肺炎病灶细粒度特征的分割与识别能力, 使模型能够在多中心胸部 CT 影像质量参差不齐的复杂环境下具有更高的稳定性与鲁棒性。

2.2 CAP 数据项查询与确定

完成数据采集、治理、结构化、标准化等环节后, 可以根据研究需求从平台筛选符合研究诊断标准、纳入排除标准的病例及能够满足统计分析的数据项。平台具备大数据交互式分析查询能力, 提供分布式结构化查询语言 (SQL) 交互式查询引擎, 支持基于内存的并行计算, 数据量支持吉字节 (GB, 1 GB = 10^9 字节) 到拍字节 (PB, 1 PB = 10^{15} 字节), 并实现秒级查询效率, 以及跨数据源的级联查询。利用平台的查询筛选功能, 选择了 255 个数据项, 分为多个数据表呈现, 其中主表 138 个数据项、检查表 12 个数据项、检验表 14 个数据项、中药表 25 个数据项、护理表 12 个数据项、医嘱表 20 个数据项。截至 2024 年 12 月, 共纳入来自 10 家中心的 5499 例符合标准的 CAP 病例。

3 基于平台的 CAP 真实世界队列的应用

3.1 研究设计

队列研究是一种广泛应用于临床研究的方法, 也是 RWS 的最常见方法之一, 包括前瞻性队列研究和回顾性队列研究。为取得较高质量循证医学证据, 避免回顾性队列研究中常见的数据质量和完整性问题, 因此采取多中心前瞻性队列研究的设计, 先确定研究方案, 再组织全省多家中心按照研究方案执行 CAP 诊疗流程并完成随访, 最后利用该平台大规模采集 CAP 患者的真实世界数据进行分析, 反映诊疗方案在实际临床中治疗 CAP 的真实疗效和安全性。

3.2 纳入、排除标准

真实世界队列研究的纳入、排除标准本质是在“真实代表性”与“研究可靠性”之间找平衡, 主动纳入临床常见的合并症、老年人及特殊人群, 提升结果外推性。具体如下: 1) 纳入标准: 出院主要诊断为 CAP 的住院患者; 年龄 18~80 岁。2) 排除标准: 合并心、脑、肝、肾、呼吸、神经、免疫、血液等系统严重疾病, 感染性疾病及恶性肿瘤者; 妊娠或哺乳期妇女。

3.3 研究分组

在队列研究中, 暴露因素是划分暴露组与非暴露组的关键依据。CAP 的中医 RWS 中, 暴露因素为 CAP 患者接受诊疗方案中的中医辨证治疗, 因此暴露组为所有使用中医辨证治疗的受试者, 非暴露组为所有未使用中医辨证治疗的 CAP 受试者。

3.4 样本量计算

如果真实世界队列研究的受试对象是有限总体, 可以不计算样本量, 将这个总体人群均纳入研究^[13]。CAP 的受试对象是有限总体, 即从研究开始后 2 年内所有参与单位的符合纳入、排除标准的 CAP 住院患者, 根据参与单位历年的患者人数, 拟定样本量为不少于 5000 例。

3.5 结局指标的选择

CAP 的中医 RWS 不仅注重对 CT、炎症指标等客观指标的观察, 也注重对临床症状等主观感受的观察, 结局指标涵盖主要结局指标、次要结局指标、经济学指标、安全性指标四个方面。

3.5.1 主要结局指标

1) 7 天临床治愈率: 对比治疗前后体温正常且症状全部消失的患者比例; 2) 21 天肺部炎症影像吸收率: 对比治疗前后胸部 CT 影像炎症吸收的患者比例。

3.5.2 次要结局指标

1) 临床稳定时间: 住院治疗后达到临床稳定状态的天数。2) 完全退热时间: 首次用药后体温降至正常范围且持续 24 h 以上的时间。3) 体温曲线下面积: 每天测量并记录患者的最高体温, 勾画体温曲线图, 计算曲线下面积。4) 7 天咳嗽症状消失率: 对比治疗前后咳嗽症状消失的患者比例。5) 4、7 天炎症指标: 对比治疗前后白细胞计数、C 反应蛋白、降钙素原的绝对值及比例的变化。

3.5.3 经济学指标

1) 成本-效果比 (CER): 是指达到某一治疗效果所花费的成本, 比值越低表明治疗方案的经济性越好。2) 增量成本-效果比 (ICER): 即比较两种治疗方案成本差值与效果差值的比值, 以判断增加的成本是否值得获得额外的效果。3) 意愿支付值 (WTP): 分析不同支付意愿下治疗方案的经济性概率。

3.5.4 安全性指标

主要为治疗前后肝功能 [丙氨酸氨基转移酶 (ALT) / 天冬氨酸氨基转移酶 (AST)]、肾功能 [血肌酐 (SCr)、尿素 (UREA)、尿素氮 (BUN)]。

4 小结

依托平台支撑, 本研究围绕 CAP 开展了中医 RWS 研究的探索。一方面, 提出真实世界数据的标准化整合与分析方法, 采用符合中医辨证论治个性化特点的数据处理与分析方式, 为应对中药处方动态变化、证候分型具有异质性等 RWS 常见数据难题提供了可行的技术路径, 为其他中医优势病种研究提供参考。另一方面, 构建了大规模、高质量的中医 RWS 队列, 整合了包含人口学特征、中医证候、中药处方、实验室检查及影像学结果等在内的系列数据, 解决了过去传统研究中数据零散, 无法量化分析的难题。此外, 基于高质量的真实世界队列数据, 采用多中心前瞻性队列研究设计, 有效规避信息偏倚, 保障数据完整性和准确性; 纳入、排除标准兼顾真实代表性与研究可靠性, 贴近临床实际, 外推性强; 以暴露因素为核心分组, 逻辑清晰、变量明确; 疗效指标覆盖主观和客观数据, 并对有效人群进行精准定位; 利用科学严谨的统计分析方法, 客观展现中医药在改善 CAP 患者的临床症状、降低炎症指标以及节约医疗资源等方面的优势。但本研究在实践过程中仍然面临问题。第一, 各家医疗机构数据标准、编码系统不尽相同, 中医数据智能化整合效率较低; 第二, 部分数据项缺失

或缺乏连续性记录, 部分疗效指标分析受限; 第三, 中医辨证论治重视随症加减、动态调整, 暂无法深入分析中医处方的加减规律和动态变化。

本研究提出的 CAP 研究模式为中医药领域的 RWS 推动了循证医学理念在中医药领域的落地与应用, 促进了中医诊疗从经验驱动向数据辅助的现代化转型。未来随着多源数据的持续积累和分析技术的不断升级, 该研究模式可以在更多的中医优势病种研究中得到实践和验证, 进而促进中医药现代化和国际化。

致谢

感谢江苏省中医药管理局、江苏省卫生健康信息中心提供政策、数据、技术支持。

参考文献

- [1] 李杨, 杜雷雷, 许飞, 等. 大数据与人工智能在医学领域的应用进展[J]. 协和医学杂志, 2023, 14(1): 184-189.
- [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] 盛松, 黄烨. 基于真实世界开展中医药研究的优势和局限性[J]. 中西医结合心脑血管病杂志, 2023, 21(2): 377-379.
- [4] 杨亦含, 胡元会, 陈婷, 等. 基于真实世界的中医临床科研大数据平台研发现状、不足及展望[J]. 中国中医基础医学杂志, 2022, 28(11): 1882-1886.
- [5] 中华中医药学会内科分会, 中华中医药学会肺系病分会, 中国民族医药学会肺病分会. 社区获得性肺炎中医诊疗指南(2018 修订版)[J]. 中医杂志, 2019, 60(4): 350-360.
- [6] 世界卫生组织. 疾病和有关健康问题的国际统计分类: 第十次修订本(ICD-10) [M]. 董景五, 主译. 北京: 人民卫生出版社, 2008.
- [7] 国家市场监督管理总局, 国家标准化管理委员会. 中医病证分类与代码[S/OL]. (2021-10-11)[2025-10-26]. <https://openstd.samr.gov.cn/bzgk/std/newGbInfo?hcno=41FD9D06E5BE4F84EA1D8D07101BED2C>.
- [8] SNOMED International. SNOMED CT (EB/OL). (2025-12-01)[2025-10-26]. <https://www.snomed.org/snomed-ct>.
- [9] 周磊, 李略, 彭玉芝, 等. 基于 PDCA 构建专病模板为导向的 CAP 患者病情严重程度评估体系[J]. 中国现代医生, 2024, 62(35): 91-93.
- [10] 国家药典委员会. 中华人民共和国药典: 一部(2025 年版)[M]. 北京: 中国医药科技出版社, 2025.
- [11] 中华人民共和国卫生部. 国家基本药物目录(2012 年版) [EB/OL]. (2013-03-15) [2025-10-26]. http://www.xayxfw.com/cn/yzgl/info_21.aspx?itemid=20837.
- [12] 江苏省药品监督管理局. 江苏省中药饮片炮制规范

(2020 年版 第一册)[EB/OL]. (2021-01-27)[2025-10-26]. https://www.jiangsu.gov.cn/art/2021/1/27/art_64797_9656482.html.

[13]翟静波, 郑文科, 王辉, 等. 真实世界研究样本量估计的统计学考虑[J]. 世界中医药, 2019, 14(12): 3123-3126.

Construction and Application of a Real-World Cohort of Community-Acquired Pneumonia Based on a Multimodal Large-Scale Traditional Chinese Medicine Big Data Platform

WANG Zhichao¹, ZHOU Xianmei¹, FENG Fanchao¹, WANG Mengqi², WANG Xin³, KANG Bin⁴, YU Xiaofan⁵, WANG Xiaoxiao¹, XIAO Lei¹, LI Juan¹, ZHANG Zhichao^{1,6}, MA Ye^{1,6}, JI Yeqing^{1,6}, TONG Xin⁷, WU Zhuoyue⁷, LIU Jia²

1. Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, 210029; 2. Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences; 3. Data Center of Chinese Medicine, China Academy of Chinese Medical Sciences; 4. School of Internet of Things, Nanjing University of Posts and Telecommunications; 5. Institute of Air-Space-Ground-Sea Communication Technology Integration, Nanjing University of Posts and Telecommunications; 6. The First Clinical Medical College, Nanjing University of Chinese Medicine; 7. School of Public Health, Nanjing Medical University

ABSTRACT This paper introduces a real-world cohort research model for community-acquired pneumonia (CAP) based on the Jiangsu Traditional Chinese Medicine (TCM) Dominant Diseases Diagnosis and Treatment Data Platform. Firstly, data cleaning is performed by standardizing diagnosis, symptoms, treatment and imaging, intelligently extracting unstructured information, and cleaning and constructing a standardized database. Secondly, for cohort establishment, CAP patients across the province are screened in accordance with CAP diagnostic criteria to build a high-quality disease-specific cohort. Lastly, in terms of protocol design, the characteristics of TCM research and the CAP disease profile are considered to determine appropriate inclusion and exclusion criteria, estimate sample size, define interventions, outcomes and economic evaluations, providing a reference for real-world TCM research on CAP.

Keywords community-acquired pneumonia; real-world study; big data platform; artificial intelligence

(收稿日期: 2025-10-25; 录用日期: 2026-01-12; 修回日期: 2026-02-28)

[编辑: 邓 媛]

广告

《中医杂志(英文版)》欢迎投稿

Journal of Traditional Chinese Medicine [《中医杂志(英文版)》] (ISSN 0255-2922, CN 11-2167/R, 双月刊) 由中华中医药学会与中国中医科学院联合主办, 是我国中医药界具有权威的中医药英文科技期刊, 现发行和交换 90 多个国家和地区。

《中医杂志(英文版)》自 1981 年创刊以来一直坚持正确的办刊宗旨, 国际影响力不断提升。2010 年入选 SCI-E (《科学引文索引》扩刊版), 2011 年获得中国中医科学院自主选题项目资助, 2013 年获得“中国科技期刊国际影响力提升计划”资助, 自 2012 年起连续被评为“中国最具国际影响力学术期刊”, 2023 年再次入选“T1 级中医药英文科技期刊目录”。本刊先后被国际知名文献数据库 MEDLINE、《化学文摘》(CA)、《医学文摘》(EM) 等收录, 是 MEDLINE 收录的第一份中国出版的中医药期刊, 是 SCI 收录的 30 余种全科医学和替代医学期刊中的中国传统医学综合性学术期刊之一。

《中医杂志(英文版)》主要栏目包括循证研究、理论研究、基础研究、临床观察、中药研究、文献研究, 思路与方法、综述等。内容以报道中医药基础、临床研究成果和科研前沿进展为主, 覆盖中医药学科各个领域。欢迎广大专家学者积极投稿。作者可直接登录本刊网站 <http://www.journaltcm.cn> 投稿。

联系邮箱: jtcmen@126.com。联系电话: 010-64089190。

联系地址: 北京市东城区东直门内南小街 16 号, 邮政编码: 100700。