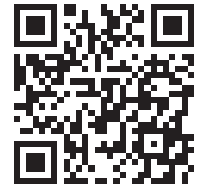


六要素医学智能体闭环管理肺结节:BAIMGPT-PN-RWE多中心真实世界研究方案



李晓明¹, 陈智鸿¹, 白春学^{2,3,4,5*}

1. 上海交通大学医学院附属新华医院呼吸与危重症医学科, 上海交通大学医学院呼吸病研究所, 上海 200092
2. 复旦大学附属中山医院呼吸与危重症医学科, 上海 200032
3. 复旦大学附属中山医院AI+肺癌防治中心, 上海 200032
4. 上海呼吸物联网医学工程技术研究中心, 上海 200032
5. 上海市呼吸病研究所, 上海 200032

[摘要] 在肺癌高危人群中筛查低剂量螺旋CT大大降低了肺癌的死亡率,却也检出了大量需要进一步风险分层和长期随访的肺结节。现有的影像人工智能多聚焦于肺结节的检出或恶性风险预测;智能提醒工具可减少部分患者的随访遗漏。但来自单一智能模型的证据尚未回答多模块医学智能体是否能改善患者层面的随访管理结局。因此,本研究的目的是明确具备感知、理解、预测、决策、执行和学习能力六要素的医学智能体能否改善肺结节的全流程随访管理和患者的临床结局。BAIMGPT-PN-RWE是一项多中心、前瞻性、基于新发现患者、活性对照的真实世界研究。计划纳入6 000名来自不同层级医疗机构的新发现4~30 mm肺结节的成年人。主要比较时间零点后7日内启动BAIMGPT六要素闭环管理随访系统与启动常规人工随访或现有物联网辅助管理随访系统的患者的随访情况。随访期为12个月(部分队列延长至24个月)。主要研究方法基于模拟目标试验,采用倾向性评分重叠加权法估计重叠人群平均处理效应;采用双重差分法分析各中心分阶段上线所带来的效应,并构建边际结构模型以控制实际使用强度带来的时依性混杂。本研究的主要结局为肺结节患者在随访期(12个月)内与指南推荐一致的随访管理的完成率。即风险分层、随访计划制定、随访执行、随访结果回传和下一步决策均在预设时间窗内完成的比率。主要安全终点包括恶性结节的漏诊、对良性结节的过度侵入性检查、高危结节的随访漏升级、以及因系统本身因素造成的随访延误。本研究评估的是真实世界中,具备六要素的医学智能体对肺结节患者随访管理的价值。基于研究数据,进一步界定六要素闭环管理医学智能体可以推广、需要限制或应当停用的条件。

[关键词] 肺结节;医学智能体;闭环管理;真实世界证据;目标试验模拟;临床人工智能

[中图分类号] R 563 **[文献标志码]** A

BAIMGPT-PN-RWE: protocol for a multicenter real-world study on six-element medical agent-based closed-loop management of pulmonary nodules

Li Xiaoming¹, Chen Zhihong¹, Bai Chunxue^{2,3,4,5*}

1. Department of Respiratory and Critical Care Medicine, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Institute of Respiratory Diseases, Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China
2. Department of Respiratory and Critical Care Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, China
3. AI+Lung Cancer Prevention and Treatment Center, Zhongshan Hospital, Fudan University, Shanghai 200032, China
4. Shanghai Respiratory Internet of Things Medical Engineering Technology Research Center, Shanghai 200032, China
5. Shanghai Institute of Respiratory Diseases, Shanghai 200032, China

[Abstract] Screening with low-dose helical computed tomography (LDCT) has substantially reduced lung cancer mortality in

[收稿日期] 2026-07-10

[接受日期] 2026-09-04

[基金项目] 四大慢病重大专项(2023ZD0506305、2024ZD0529300),国家自然科学基金(82570143、82170110),上海市科学技术委员会项目基金(20DZ2254400、21DZ2200600、20DZ2261200),上海市市级科技重大专项(ZD2021CY001)和上海市临床重点专科建设项目(shslczdk02201)。Supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0506305, 2024ZD0529300), the National Natural Science Foundation of China (82570143 and 82170110), the Science and Technology Commission of Shanghai Municipality (20DZ2254400, 21DZ2200600, and 20DZ2261200), the Shanghai Municipal Science and Technology Major Project (ZD2021CY001), and the Shanghai Municipal Key Clinical Specialty (shslczdk02201)。

[作者简介] 李晓明,博士,主任医师;E-mail: lixiaoming13734@126.com

*通信作者(Corresponding author). 白春学,博士,主任医师、教授;E-mail: cxbai@fudan.edu.cn

high-risk populations, but it has also led to the detection of large numbers of pulmonary nodules that require further risk stratification and long-term follow-up. Existing imaging artificial intelligence (AI) tools have focused primarily on nodule detection or prediction of malignancy risk, and smart reminder tools can reduce missed follow-up in some patients. However, evidence from single AI models has not yet answered whether multi-module medical agents can improve patient-level follow-up outcomes. Therefore, this study aims to determine whether a medical agent equipped with six capabilities—perception, comprehension, prediction, decision-making, execution, and learning—can improve whole-process follow-up management of pulmonary nodules and patient clinical outcomes. BAIMGPT-PN-RWE is a multicenter, prospective, active-controlled real-world study enrolling patients with newly detected nodules. We plan to enroll 6,000 adults with newly detected pulmonary nodules measuring 4 – 30 mm from healthcare institutions at different tiers. The primary comparison is between patients who initiate the BAIMPT six-element closed-loop follow-up management system within 7 days after time zero and those who initiate routine manual follow-up or an existing Internet of Things (IoT) - assisted management system. The follow-up period is 12 months (extended to 24 months in a subset of the cohort). The primary analysis emulates a target trial: propensity score overlap weighting will be used to estimate the average treatment effect in the overlap population; a difference-in-differences approach will be used to assess the effect of the staggered rollout across centers; and marginal structural models will be constructed to control for time-dependent confounding arising from the actual intensity of system use. The primary outcome is the completion rate of guideline-concordant follow-up management during the 12-month follow-up period, i.e., the proportion of patients in whom risk stratification, follow-up planning, follow-up execution, reporting of follow-up results, and next-step decision-making are all completed within prespecified time windows. Key safety endpoints include missed diagnosis of malignant nodules, excessive invasive examinations of benign nodules, failure to escalate follow-up for high-risk nodules, and follow-up delays attributable to the system itself. This study evaluates, in the real world, the value of a six-capability medical agent for follow-up management of patients with pulmonary nodules. Based on the study data, we will further delineate the conditions under which six-element closed-loop medical agents can be generalized, should be restricted, or should be discontinued.

[Key Words] pulmonary nodules; medical agents; closed-loop management; real-world evidence; target trial emulation; clinical artificial intelligence

NLST 和 NELSON 研究^[1-2]表明,在肺癌高危人群中筛查低剂量螺旋 CT 大大降低了肺癌死亡率。但是在临床获益的同时检出了大量需要进一步风险分层的肺结节,并且加重了随访负担。Fleischner Society、BTS、Lung-RADS 及亚洲肺结节专家共识^[3-6]均推荐综合考虑肺结节大小、密度、形态、生长速度和患者风险,选择继续观察、复查胸部 CT、完善 PET-CT、活检或手术。然而,肺结节管理的核心不是最大化检查或手术,而是在良性结节过度检查、恶性结节漏诊和跨机构随访脱落之间取得平衡。

深度学习模型和基于临床—影像的风险预测模型已被证实在肺结节检出和恶性风险预测中具备较好的性能;自动登记和智能提醒系统可以减少部分患者的随访失败。但上述模型的准确性不会自动转化为患者的临床获益^[7-9]。研究^[10-12]表明,涵盖更多个管理环节的人工智能较单一智能模型在改善患者随访方面更具潜力。然而,多数工具仍局限于“发现—提醒”环节,缺少风险预测、管理失败预测、决策、任务执行和结局反馈等模块。

研究版 BAIMGPT 将上述能力组织为“感知—理解—预测—决策—执行—学习”的闭环式六要素(图 1)^[13-14]。系统模块在肺结节物联网辅助评估管理路径的基础上扩展至任务执行与结局反馈^[15],并

预设人类监督、分级授权、全程留痕和版本控制等环节。拟明确真实临床环境中,与常规人工随访或物联网辅助管理随访系统相比,新诊断的肺结节患者启动 BAIMGPT 六要素管理系统能否提高其在 12 个月内与指南推荐一致的闭环式随访管理的完成率,同时不增加恶性结节漏诊和良性结节过度侵入性检查的风险。

1 方法

1.1 研究设计与报告框架

本研究为多中心、前瞻性、基于新发现人群,活性对照的观察性队列,嵌入常规临床工作流程。研究利用系统分阶段上线自然形成的管理策略差异作为分组依据,而非由研究者人为分配干预措施。主要分析在患者层面采用目标试验模拟(target trial emulation)框架;中心分阶段上线作为准实验设计,进行互补性因果推断;实际使用强度作为时间变化暴露进行机制性分析。研究报告将依据 STROBE、RECORD 和 TARGET,并参照 DECIDE-AI、TRIPOD+AI 及 MI-CLAIM-GEN,披露临床 AI 系统、模型版本、工作流交互和错误事件^[16-22]。

研究设计遵循循证医学 PICO 原则,即按研究对象、策略、对照和结局四个维度对研究进行规范性

限定。在符合入选条件且两种随访策略均适用的患者中,观察时间零点后 7 日内启动 BAIMGPT 随访系统与启动常规随访管理系统对 12 个月内肺结节闭环随访管理完成率的影响。主要目标效应量

estimand 为倾向评分重叠人群中的平均处理效应 (average treatment effect in the overlap population, ATO)(表 1)。

表 1 拟模拟的目标试验方案

目标试验要素	预设方案
研究人群	新发现 4~30 mm 肺结节、尚无明确良恶性诊断且需要进一步评估或随访的成年人
随访策略 A	时间零点后 7 日内启动研究版 BAIMGPT 六要素闭环管理系统
随访策略 B	启动常规人工随访、肺结节门诊或现有物联网辅助管理随访系统,但不具备完整六要素
时间零点	首次符合条件的胸部 CT 完成、入选条件确认、管理路径确定与随访开始同日对齐
宽限期	7 日;通过克隆—删失—加权或等价策略处理宽限期引入的偏倚
主要随访	12 个月;部分患者延长至 24 个月
主要结局	12 个月内与指南推荐一致的闭环随访管理完成率
因果对比	策略启动的意向治疗分析效应;另报告按方案及实际使用强度分析
主要目标效应量	重叠人群平均处理效应(ATO);报告绝对风险差、风险比和 95% CI
结局判定	对随访策略设盲的独立结局委员会,使用预先制定的判定章程

1.2 研究场所与研究对象

研究中心覆盖不同地区和医疗水平。计划纳入 30~40 家三级医院、区域肺结节中心、地市级医院、基层筛查机构和体检中心。所有中心采用统一的数据字典、终点判定规则和安全事件上报流程。纳入标准:(1)年龄≥18 岁;胸部 CT 新发现至少 1 个直径 4~30 mm 实性、部分实性或磨玻璃结节;(2)尚无明确良恶性病理诊断,临床需要风险评估、复查胸部 CT、PET-CT、MDT、活检或外科评估;(3)可获得原始 DICOM 或结构化影像信息、影像报告和基本临床资料;(4)同意按伦理委员会批准的方案使用去标识化的临床资料、影像、平台日志和随访数据。排除标准:(1)已明确的活动性肺癌、其他肿瘤肺转移,或入组前已针对目标结节实施活检、消融或手术;(2)影像具有明确的良性特征且无需随访,或急性感染已行抗感染治疗;(3)无法确定时间零点,或无法获得任何后续结局资料。允许纳入既往肺癌或肺纤维化的患者,但需进入预设的扩展队列并单独报告。多发结节患者以预设目标结节作为结节层面分析的主要单位,同时保留患者层面的管理结局分析。

1.3 随访管理策略、暴露与污染

BAIMGPT 组患者在时间零点后 7 日内启动六要素闭环管理系统,即完成策略激活,并至少完成 1 次可追溯的风险评估和随访任务。对照组患者则接受常规人工随访、专病门诊或物联网辅助管理随访。主要分析按初始策略分组;之后患者无论是否

继续使用平台,均不改变其原有分组。系统首次激活、数据接入、医生审核、患者交互、任务生成与完成、人工接管均记录原始时间。知识库和模型更新记录版本号。

对照组各中心医生接触 BAIMGPT 系统输出的随访建议将视为污染事件。研究将记录跨路径咨询、共享账号、系统导出建议和中心切换等事件。通过研究中心及医生的层级分析、按方案分析及污染阈值敏感性分析评估暴露和污染事件对结局的影响。

1.4 BAIMGPT 六要素闭环管理

研究版 BAIMGPT 是受医生控制的临床决策与任务编排系统,包括感知—理解—预测—决策—执行—学习六个模块(表 2)。BAIMGPT 不独立确诊肺癌,不自主决定活检、手术、消融或放疗等诊疗行为,也不取消指南推荐的必要随访。最终医学决策由具备相应资质的临床医生作出。

1.5 结局

1.5.1 主要结局 本研究的主要结局为肺结节患者在 12 个月内与指南推荐一致的闭环随访管理完成率。患者需同时满足:(1)时间零点后 7 日内生成明确、可追溯的风险分类与随访管理计划;(2)在肺结节指南或专家委员会的推荐下完成相应的胸部 CT 复查、PET-CT、MDT、肺结节活检或其他计划;(3)随访结果上传至系统并由医生确认;(4)结果产生后 14 日内生成继续随访、终止随访或进一步诊疗决定;(5)12 个月内无重大随访管理偏差(表 3)。

六要素医学智能体闭环及其可检验机制



图 1 六要素闭环的作用路径与管理边界

表 2 六要素干预、机制与可统计指标

要素	研究版功能	主要机制	关键量化指标
感知	整合胸部 CT、结构化报告、既往影像、临床风险、患者偏好;区分信息缺失与阴性	减少信息遗漏	数据完整率、历史影像匹配率、错配率
理解	肺结节特征、具体临床情况和适用指南;识别指南冲突与信息缺失	提高状态解释一致性	专家一致率、指南推荐匹配率、指南推荐冲突率
预测	分别评估肺结节的恶性、进展和管理失败风险;在研究对象中校准	提前识别高风险	判别度、校准度、净获益
决策	结合肺结节的恶性风险、进展风险、手术风险、患者偏好和本地资源生成患者可获得的候选随访方案	提高与指南推荐的一致性	与指南推荐的一致率、医生对系统输出决策的修改/否定率
执行	输出随访建议、患者教育、辅助预约、智能提醒、生成 MDT 或转诊工单并追踪完成	缩短行动延迟	任务完成率、节点时间、无效警报率
学习	更新患者状态并进行校准;系统更新需独立验证、并提交最新版伦理审批	维持适配性	版本记录、校准频率、回溯与变更统计

若患者提前获得明确病理诊断、结节消失,或已按指南完成全部管理路径,则视为闭环终止。重大随访管理偏差包括恶性结节的漏诊或对良性结节的过度侵入性检查。上述两类安全性事件的定义、允许时间窗及其与系统的相关性,均由独立结局委员会进行判定。

表 3 主要结局、机制、安全性及公平性结局

层级	结局	时间点	效应量/描述指标
主要	与指南推荐一致性的闭环随访管理完成率	12 个月	调整后风险差、风险比和 95%CI
临床安全	恶性结节漏诊;良性结节过度侵入性检查	12 个月;扩展随访至 24 个月	累积发生率、风险差及风险比
机制	首次识别到任务至任务完成的各节点耗时	全程	中位数、限制平均时间及中介效应
实施	医生投入时间、对系统决策的采纳/否决率、无效警报率、额外成本	12 个月	均值差、率比、增量成本效果比
AI 安全	有害建议、高危结节漏升级、患者或结节信息错误匹配、系统停机延误	全程	每 1 000 次建议或每 1 000 位受试者的日发生率
公平性	智能系统的可及性、闭环完成率与错误使用率的绝对差异	12 个月	交互作用、标准化绝对差异

1.6 数据来源与质量控制

具有患者标识的原始数据来源于医院信息系统、电子病历、PACS/DICOM、影像 AI、BAIMGPT 日志、肺结节随访平台、病理与手术资料、肿瘤登记、患者结局报告和区域转诊记录。肺结节数据与患

者数据分别建模,保留原始记录时间,不以人工汇总日期替代。

研究启动前统一各研究中心的数据字典,逻辑核查、重复记录合并和跨院结局回收规则。独立委员会对关键变量进行源数据抽查;主要结局、恶性

结节漏诊和良性结节过度侵入性检查由独立委员会盲法复核。系统日志保存信息的输入来源、输出、置信区间或不确定性提示、所引指南、医生处置、人工接管、知识库和模型版本等内容。

1.7 样本量

假设对照组 12 个月闭环随访完成率为 60%，以试验组的完成率较对照组提高 5% 设为最小具有临床意义的差异。在双侧 $\alpha=0.05$ 、把握度 90%、两组样本量相等的条件下，总计约需 3 936 例患者。根据既往多中心研究的中心内相关系数(ICC)估计，设定中心聚集和策略异质性造成的设计效应为 1.20，以及约 15% 的失访率或关键数据缺失，最终确定目标样本量为 5 557 例。研究计划纳入 6 000 例受试者(每组约 3 000 例)，尽可能将至少 1 500 例受

试者的随访时间延长至 24 个月。最终样本量将在不接触结局组间差异的前提下，依据先期对照组发生率、中心内相关系数和有效加权样本量进行盲态复核。本研究不具备检验肺癌死亡率差异的统计效能。

1.8 统计分析和偏倚控制

主要分析遵循预先制定的统计分析计划。连续变量和分类变量采用相应的描述统计(表 4)。基线特征的组建平衡采用标准化差异评价。混杂因素的筛选根据有向无环图并结合临床实际确定，包括人口学、吸烟史与肿瘤风险、基线可获得的结节特征及既往生长信息、肺部背景、合并症、医疗可及性、数字素养、中心层级、日历时间和系统版本。

表 4 主要分析与偏倚控制

分析	方法	目的/关键假设
主要效果	采用倾向性评分重叠权重法进行加权，通过加权广义线性模型或标准化估计风险差和风险比；标准误采用考虑中心内聚集性的稳健标准误	估计 ATO；检验可交换性、阳性与模型设定
宽限期	克隆—删失—加权法或等价方法；并比较 3 日、7 日和 14 日定义	避免策略启动宽限期造成不死时间偏倚
分期上线	采用适用于分期处理的组别—时间平均处理效应与事件研究；中心和日历时间效应纳入模型	利用上线时序提供准实验互证；检验上线前趋势并避免仅依赖传统双向固定效应
实际使用	逆概率治疗加权的边际结构模型	处理交互、采纳与任务执行的时间变化混杂
时间结局	累积发生函数、限制平均生存时间和原因特异风险模型	处理竞争事件并减少比例风险假设依赖
缺失数据	描述缺失机制；多重插补；跨院失联采用逆概率删失加权	比较完整案例、插补及极端情景结果
稳健性	负对照、E-value、定量偏倚分析、中心留一法、污染与版本分层分析	评估未测量混杂和单中心/单版本驱动

主要终点报告绝对风险差、风险比和 95%CI；不以单一 P 值代替效应量及其不确定性。关键次要终点按预设层级或多重性策略解释，其他分析标注为探索性。异质性分析预设年龄、性别、吸烟状态、筛查/偶然发现、实性/亚实性结节、中心层级、地区和数字素养；亚组间差异的判定依据交互项检验及其对应的绝对效应差，而非仅看单个亚组是否达到名义上的统计学显著。

核心机制分析主要检验“识别异常—医生审核—患者通知—随访预约—检查完成—检查结果回传—下一步决策”的节点时间是否缩短，分析结果仅用于解释随访流程改善，不代表患者层面临床结局。

1.9 实施与公平性评价

采用 BAIMGPT 系统的覆盖度、系统输出建议的

采纳度、系统使用的一致性、维护成本等指标评价研究的具体实施过程。患者端记录激活机会、是否持续使用、有无通过电话或线下途径替代和家属协助使用平台系统情况；临床端记录医生的使用时间、对系统建议的采纳/修改/否决、无效警报与人工接管。无法使用智能手机或数字素养较低的受试者必须通过电话、门诊或家属协助等途径进行随访。

公平性评价主要比较不同年龄、性别、地区、医院层级、受教育水平等因素的受试者完成闭环随访管理的比率。报告由上述因素导致的随访差异和交互作用对结局的影响。进一步评估缺失数据、医生未响应和模型失准是否集中于受教育水平较低的受试者人群。

1.10 安全、权限与系统管理

所有由系统输出的随访建议均标注证据来源、适用条件和不确定性。研究期间禁止未记录的在线权重更新或决策阈值改变。基于肺结节恶性风险的系统分级授权与人类监督机制见表 5。任何模型、知识库、提示词、规则或接口的更新均需经过版

本锁定、离线验证和审批流程,并确保更新记录可回溯。独立安全委员会定期审查高风险错误和系统停机事件。WHO 关于健康 AI 管理的原则将作为研究过程中隐私保护、算法透明、责任追溯与人类监督的基本框架^[23]。

表 5 分级授权与人类监督

风险层级	任务示例	系统权限与控制
低风险	健康教育、资料补充、按既定计划提醒	可自动执行;必须留痕并允许患者退出
中低风险	预约胸部 CT、提醒上传既往影像	按医生预设随访计划执行;患者确认
中风险	建议 MDT、PET-CT 或专科转诊	医生审核后执行;记录采纳或否决建议,并给出理由
高风险	建议活检、手术、消融或放疗	必须由相应专科医生决定;系统不得直接下达

出现可能与系统相关的严重漏诊、未经医生确认的侵入性检查、重复高风险错误、患者或结节错配、严重隐私泄露、低受教育水平人群错误率异常升高或版本不可追溯时,应立即触发暂停、根因分析和独立复核机制。恢复条件须预先写入安全章程。

1.11 伦理、注册与数据共享

研究将遵循《赫尔辛基宣言》的数据保护法规,由牵头单位和各中心伦理委员会审批。知情同意方式、常规诊疗数据的使用、患者端交互及跨机构数据链接均按本地伦理委员会要求执行。研究将在首例患者入组前完成公开注册。研究方案、终点判定章程、统计学方法和关键修订进行存档并记录原始时间。

本方案不报道单个受试者的研究数据。研究完成后,将在伦理委员会批准、数据安全要求和参与机构政策允许范围内通过受控访问的方式共享去标识的原始数据集、数据字典和分析代码。无法共享的数据及理由将在研究论文中明确说明。

2 讨 论

本研究将评价肺结节智能随访管理模型从“性能”到“连续管理效果”的关键转化环节。以一个主要终点、一条机制链和两类安全边界为中心,检验肺结节患者全流程闭环随访管理的完成情况。节点时间缩短用于解释六要素智能体改善肺结节患者全流程随访管理和临床结局的潜在机制。诊断延误和过度侵入性检查界定随访效率提升是否以患者伤害为代价。

研究的主要优势是采用新发患者活性对照设计、严格统一时间零点、多种因果推断方法互证、以患者层面的临床结局为主要评价指标,以及系统层

面的可审计性和公平性评价。目标试验模拟框架使观察性研究问题更贴近临床决策的实际情境,但无法完全消除未测量混杂的影响;分期上线分析仍依赖无预期效应、平行趋势和无同期差异性冲击等假设;实际使用分析易受健康使用者偏倚和时间变化混杂影响。跨中心结局缺失、中心实施能力差异、干预措施交叉污染、亚实性结节较长的自然史和系统版本演进也可能一定程度上限制结果的解释。为此,本研究将通过预设负对照、定量偏倚分析、事件研究法、版本分层分析和 24 个月扩展随访等手段对上述潜在偏倚进行量化评估。

若主要分析、分期上线分析和敏感性分析方向一致,且安全性与公平性结果可接受,则可认为 BAIMGPT 六要素闭环管理系统与更高的闭环完成率相关。但不能据此认定其可降低肺癌死亡率、替代临床医生或适用于所有肺结节。若仅缩短医生响应时间,那么研究结论应局限于改进肺结节随访效率;若发生恶性结节漏诊、良性结节过度侵入性检查或群体差异增加,则应暂停扩展,并重新校准决策阈值、权限和实施路径。

综上所述,六要素医学智能体并非孤立算法,而是一种可统计和检验的管理策略。本研究将通过目标试验模拟、准实验互证和明确的安全性边界评估六要素医学智能体对真实世界肺结节患者闭环随访管理的价值,并明确其可以推广、需要限制或应当停止的条件。

伦理声明 无。

利益冲突 所有作者声明不存在利益冲突。

作者贡献 李晓明:素材整理、文章撰写;陈智鸿:素材提供、稿件审核;白春学:设计研究框架、文章审核。

参考文献

- [1] National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening [J]. *N Engl J Med*, 2011, 365(5): 395-409.
- [2] de Koning H J, van der Aalst C M, de Jong P A, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial [J]. *N Engl J Med*, 2020, 382(6): 503-513.
- [3] MacMahon H, Naidich D P, Goo J M, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the fleischner society 2017 [J]. *Radiology*, 2017, 284(1): 228-243.
- [4] Callister M E, Baldwin D R, Akram A R, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules [J]. *Thorax*, 2015, 70 (Suppl 2): ii1-ii54.
- [5] Bai C X, Choi C M, Chu C M, et al. Evaluation of pulmonary nodules: clinical practice consensus guidelines for Asia [J]. *Chest*, 2016, 150(4): 877-893.
- [6] American College of Radiology. Lung-RADS v2022: assessment categories and management recommendations [EB/OL]. [2026-07-23]. <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads>.
- [7] McWilliams A, Tammemagi M C, Mayo J R, et al. Probability of cancer in pulmonary nodules detected on first screening CT [J]. *N Engl J Med*, 2013, 369(10): 910-919.
- [8] Ardila D, Kiraly A P, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography [J]. *Nat Med*, 2019, 25(6): 954-961.
- [9] Mikhael P G, Wohlwend J, Yala A, et al. Sybil: a validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography [J]. *J Clin Oncol*, 2023, 41(12): 2191-2200.
- [10] Hammer M M, Kapoor N, Desai S P, et al. Adoption of a closed-loop communication tool to establish and execute a collaborative follow-up plan for incidental pulmonary nodules [J]. *AJR Am J Roentgenol*, 2019, 212(5): 1077-1081.
- [11] Shelver J, Wendt C H, McClure M, et al. Effect of an automated tracking registry on the rate of tracking failure in incidental pulmonary nodules [J]. *J Am Coll Radiol*, 2017, 14 (6): 773-777.
- [12] Aunger J, Yip K P, Dosanjh K, et al. Interventions to improve adherence to clinical guidelines for the management and follow-up of pulmonary nodules: a systematic review [J]. *Chest*, 2025, 168(1): 248-268.
- [13] 白春学. 肺结节专家——BAIMGPT白皮书 [J]. 元宇宙医学, 2025, 2(2): 55-64.
- [14] 白春学. AI智能体赋能肺结节诊治研发进展及展望 [J]. 元宇宙医学, 2026, 3(2): 83-90.
- [15] 中国物联网辅助评估管理肺结节专家组. 物联网辅助评估管理肺结节中国专家共识 [J]. 国际呼吸杂志, 2022, 42(1): 5-12.
- [16] von Elm E, Altman D G, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies [J]. *Lancet*, 2007, 370(9596): 1453-1457.
- [17] Benchimol E I, Smeeth L, Guttman A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement [J]. *PLoS Med*, 2015, 12 (10): e1001885.
- [18] Hernán M A, Robins J M. Using big data to emulate a target trial when a randomized trial is not available: table 1 [J]. *Am J Epidemiol*, 2016, 183(8): 758-764.
- [19] Cashin A G, Hansford H J, Hernán M A, et al. Transparent reporting of observational studies emulating a target trial—the TARGET statement [J]. *JAMA*, 2025, 334(12): 1084-1093.
- [20] Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI [J]. *BMJ*, 2022, 377: e070904.
- [21] Collins G S, Moons K G M, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods [J]. *BMJ*, 2024, 385: e078378.
- [22] Miao B Y, Chen I Y, Williams C Y K, et al. The MI-CLAIM-GEN checklist for generative artificial intelligence in health [J]. *Nat Med*, 2025, 31(5): 1394-1398.
- [23] World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models [R]. Geneva: WHO, 2024.

引用本文

李晓明, 陈智鸿, 白春学. 六要素医学智能体闭环管理肺结节: BAIMGPT-PN-RWE 多中心真实世界研究方案 [J]. 元宇宙医学, 2026, 3(3): 200-206.

Li X M, Chen Z H, Bai C X. BAIMGPT-PN-RWE: protocol for a multicenter real-world study on six-element medical agent-based closed-loop management of pulmonary nodules [J]. *Metaverse Med*, 2026, 3(3): 200-206.