

三方共建共享下的肺结节/肺癌专病全程管理数据库构建

罗艺婷 杨先碧 杜春霖 张 红 柴 琪 邓 懿 周永召

(四川大学华西医院全程管理中心 成都 610041)

〔摘要〕 **目的/意义** 构建肺结节/肺癌专病全程管理数据库, 实现全病程管理与科研支撑, 助力肺结节/肺癌诊疗技术创新和发展。**方法/过程** 采用临床团队、全程管理中心、信息中心三方共建共享模式, 按疾病诊治发展阶段将数据库拆分为 4 个子数据库, 整合前瞻性 with 回顾性数据, 构建专病全程管理数据库。**结果/结论** 该数据库 2022 年建立, 截至 2025 年 7 月 1 日, 已纳入患者 36 262 例。突破传统科研项目驱动模式, 实现从肺结节发现到治疗、康复全生命周期信息追踪, 为专病数据库建设提供新范式。

〔关键词〕 专病数据库; 全程管理; 肺结节; 肺癌

〔中图分类号〕 R-058 **〔文献标识码〕** A **〔DOI〕** 10.3969/j.issn.1673-6036.2025.09.012

Construction of a Full-process Management Database for Pulmonary Nodules/Lung Cancer Specialized Diseases under Tripartite Co-construction and Sharing

LUO Yiting, YANG Xianbi, DU Chunlin, ZHANG Hong, CHAI Qi, DENG Yi, ZHOU Yongzhao

Integrated Care Management Center, West China Hospital, Sichuan University, Chengdu 610041, China

〔Abstract〕 **Purpose/Significance** To construct a database for the full-process management of pulmonary nodules/lung cancer, to achieve full-process management and scientific research support, in order to promote technological innovation and development in the diagnosis and treatment of pulmonary nodules/lung cancer. **Method/Process** A tripartite co-construction and sharing model involving the clinical team, integrated care management center, and information center is adopted. According to the development stages of disease diagnosis and treatment, the database is divided into four sub-databases, and prospective and retrospective data are integrated to construct a database for the full-process management of specialized diseases. **Result/Conclusion** Since its establishment in 2022, as of July 1, 2025, this database has included 36 262 patients. The database breaks through the traditional research project driven model and achieves full life-cycle information tracking from the discovery of pulmonary nodules to treatment and rehabilitation, providing a new paradigm for the construction of specialized disease databases.

〔Keywords〕 specialized disease database; full-process management; pulmonary nodule; lung cancer

〔修回日期〕 2025-08-03

〔作者简介〕 罗艺婷, 研究实习员, 发表论文 2 篇; 通信作者: 周永召, 博士, 副研究员。

〔基金项目〕 国家科技重大专项 (项目编号: 2024ZD0524300)。

1 引言

随着数智技术在医疗领域的深入应用,健康医疗大数据成为推动医疗行业变革的核心要素。2016 年 6 月发布的《关于促进和规范健康医疗大数据应用发展的指导意见》提出,要加强临床和科研数据资源整合共享。2021 年发布的《“十四五”优质高效医疗卫生服务体系建设实施方案》指出,要聚焦重点病种和专科,加强信息化和科研平台建设,推动临床科研成果转化,加快解决医学数据等领域“卡脖子”问题。《“健康中国 2030”规划纲要》明确提出,要加强健康医疗大数据应用体系建设,推进基于区域人口健康信息平台的医疗健康大数据开放共享、深度挖掘和广泛应用。目前,我国卫生健康信息互联互通已从 1.0 阶段进入 2.0 阶段,这一突破为健康医疗大数据互联互通指明新方向^[1]。专病数据库作为健康医疗大数据重要组成部分,是推动其应用体系建设的关键动力。

在我国,2024 年肺癌新发病例约 68.1 万例,死亡病例约 47.8 万例^[2]。肺结节是肺癌重要前期表现,其良恶性鉴别困难。同时,肺结节形成机制、肺癌发生发展过程仍存在诸多未知,建设专病队列开展真实世界研究尤为重要。专病队列研究数据具有大样本、长周期、多源异构等特征,因此专病数据库须满足覆盖肺结节/肺癌全生命周期、记录全病程多源异构数据的要求。基于此,四川大学华西医院全程管理中心联合临床科室和信息中心构建肺结节/肺癌专病全程管理数据库,助力肺结节/肺癌诊疗领域技术创新和发展。

2 专病数据库建设现状

2.1 国外专病数据库建设现状

国外专病数据库建设起步较早,共享开放程度高,在多领域取得显著成果。美国国家癌症研究所的癌症监测、流行病学和结果(surveillance, epidemiology, and end results, SEER)数据库建立于 1973 年,覆盖美国 17 个州县,约 30% 美国人口,

包含肿瘤登记及医保数据,为肿瘤研究提供大量高价值病例数据^[3]。Wingo P A 等^[4]利用 SEER 数据库分析癌症整体情况,首次总结肺癌与吸烟的关系,为癌症预防和控制提供了重要依据。Qie S 等^[5]利用 SEER 数据库,采用多种机器学习算法,发现 XGBoost 模型在预测小细胞肺癌骨转移方面表现最佳。日本的国家临床数据库(national clinical database, NCD)成立于 2010 年,是全国性、基于 Web 的外科手术数据输入系统,覆盖日本大部分手术病例,与癌症登记数据库、医疗保险数据库互通,数据标准化程度高^[6]。Baba H^[7]指出 NCD 数据库在评估和改进外科手术质量方面具有重要作用。英国生物样本库(UK Biobank, UKB)于 2002 年被提出,2006 年开始收集数据,包含癌症、心脏病、糖尿病及特定精神疾病等遗传、生活方式及健康数据^[8]。Deng Y T 等^[9]借助 UKB 蛋白质组学数据,系统描绘健康状态下血浆蛋白分布特征,及其在多种常见疾病中的变化规律和潜在生物学意义。

国外专病数据库历经多年发展,普遍具有样本量大、覆盖时间长、来源多元且针对性强的特点。然而,国外专病数据库也存在数据整合困难、数据质量参差不齐等问题。如美国 SEER 数据库,由于医疗服务分散,患者在社区医疗机构和专科医疗机构中的就诊记录难以实现电子衔接,造成家族病史等重要信息的缺失。

2.2 国内专病数据库建设现状

医疗大数据蓬勃发展,国内专病数据库建设稳步推进。我国专病数据库主要分为两种形式:国家层面推动、多中心合作共建的数据库;单中心医院建设的数据库。国家层面,国家卫生健康委员会医院管理研究所以大数据为支撑,基于多中心乳腺癌注册和全周期随访体系建立中国乳腺癌标准数据库^[10]。中国胃肠肿瘤外科联盟建立胃癌专病标准化数据库,覆盖 30 个省份,形成南北均衡的多中心协作网络,包含手术类型、并发症、病理分期等 44 项核心变量^[11-12]。2017 年 9 月中华医学会外科学分会结直肠外科学组设计并建立中国结直肠癌手术病例登记数据库(Chinese colorectal cancer surgery database,

CCCD), 录入 2018 年 3 月—2019 年 10 月全国多中心结直肠癌手术病例^[13]。单中心医院层面, 四川大学华西医院结直肠外科于 2007 年以真实世界研究为思路建立华西肠癌数据库, 患者信息包含 346 个数据特征, 涵盖入院、新辅助治疗、手术、随访等全流程环节^[14]。北京协和医院结直肠专业组于 2016 年 1 月正式建立结肠癌专病数据库, 该数据库可与医院信息系统(hospital information system, HIS)关联, 由专职科研助理管理, 截至 2022 年已完成近 2 000 例手术治疗结肠癌患者的数据录入及随访^[15]。中国中医科学院广安门医院依托中医药循证医学研究平台, 建立规范、科学的肺结节中医专病库, 收录 2017 年 1 月—2023 年 6 月共 43 976 例患者的医疗数据^[16]。北京大学肿瘤医院基于信息化病历系统和多学科会诊模式, 建立胃癌专病数据库^[17]。天津市宝坻区人民医院以心肺血管疾病患者临床诊疗数据为基础, 采用抽取、转换和加载(extraction - transformation - loading, ETL)方式实现多源异构数据抽取与转换、数据治理、质控等功能, 构建心肺血管专病数据库, 并于 2023 年 7 月投入使用^[18]。

在国家政策支持和技术创新的基础上, 我国专病数据库建设取得一定进展, 部分医疗机构和相关组织以队列研究和单病种真实世界研究为驱动, 建立专病数据库。然而, 我国专病数据库建设仍存在诸多不足。例如, 各数据库分散建设, 缺乏协调统一, 导致数据系统性和连贯性欠佳, 难以从宏观层面为公共卫生管理和政策制定提供全面、精准支持; 数据开放共享程度低, 多数数据仅在单一区域或医疗集团内部流通, 跨机构、跨地区共享案例稀少; 数据准确性和规范性有待提高, 数据采集、录入缺乏严格标准和审核机制; 不同医疗机构采用的技术架构和数据标准不同, 数据交换和整合困难重重, 严重制约数据利用效率和价值发挥。

3 专病全程管理数据库建设模式及流程

3.1 专病全程管理数据库建设模式

肺结节/肺癌临床研究数据由病历、病理、检查、检验等多维度、多类型、数量庞大的数据构成, 且生命周期长, 建立专病全程管理数据库意义重大。然而, 肺结节/肺癌专病全程管理数据库建设存在以下难点: 一是医院内部各系统存在信息壁垒, 数据分散, 形成数据孤岛; 二是临床原始数据结构化程度较低, 可用性较差; 三是数据管理缺乏标准, 数据库后期无专人维护, 易形成“烂尾库”; 四是肺结节发展成肺癌的周期较长, 患者易失访导致数据缺失。

为解决建设难点, 四川大学华西医院提出三方共建共享的专病全程管理模式, 由临床团队、全程管理中心、信息中心共同建设数据库并共享成果。临床团队专注医疗与科研, 明确患者纳排标准和数据采集需求, 协助全程管理中心为个性化采集字段自建数据标准。全程管理中心组建于 2022 年, 联合多科室实行“医护管”一体化全程管理模式, 设专病管理师、专病数据库管理师、专病推广师 3 个办公室, 分别负责专病患者管理、专病数据管理、专病项目管理^[19]。全程管理中心为各科室搭建沟通桥梁, 负责建设与维护专病数据库。其中专病管理师负责患者全生命周期随访, 采集院外及随访数据; 专病数据库管理师负责数据结构化处理及挖掘, 定期维护数据库, 搭建可持续发展数据库。信息中心专注信息治理和智慧化建设, 根据需求及数据标准开放院内数据接口, 借助数据结构化治理成果从源头对数据进行前结构化治理。

3.2 专病全程管理数据库建设流程 (图 1)

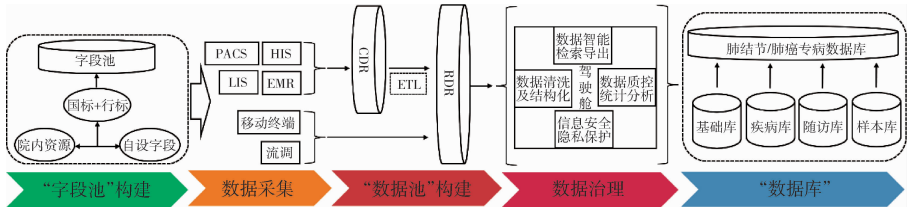


图 1 肺结节/肺癌专病数据库建设流程

在三方共建共享模式下,针对肺结节/肺癌疾病特点,依据疾病诊治发展阶段,将肺结节/肺癌专病数据库拆分为肺结节全程管理、肺部穿刺活检、肺部手术治疗、肺癌非手术治疗 4 个子数据库。子数据库间互联互通,有通用数据字段和各子库特色数据字段,患者可从多通道入组,肺结节/肺癌专病数据库可追踪不同阶段信息,形成专病链。构建专病数据库须整合多源异构数据,确定疾病标准化数据集,建立数据存储或录入逻辑,具体数据库建设流程大致分为“字段池”构建、数据采集、“数据池”构建、数据治理。

3.2.1 “字段池”构建 “字段池”构建是建设数据库的起始环节,为整个数据库内容架构奠定基础。该流程核心是全面梳理并确定与专病相关的各类数据字段,形成丰富且具有针对性的字段集合。临床团队凭借专业知识,向全程管理中心提供专病相关数据字段;全程管理中心将其分类整理,形成病例报告表(case report form, CRF)。按数据性质和用途,字段可划分为患者基本信息、临床诊疗信息、检查检验信息、随访信息等类别。查阅国家标准和行业标准制定标准化数据集;对暂无行业标准和国家标准的个性化字段,由临床团队提出采集标准,由全程管理中心整理成数据集。专病数据库“字段池”数据分两类,一是院内各系统已有数据字段;二是患者院外数据字段,如是否在外院治疗、生活自理能力等。全程管理中心与临床团队反复讨论,最终形成科学、全面、合理的“字段池”,为后续数据采集提供明确指导。

3.2.2 数据采集 数据采集是将“字段池”数据字段转化为实际数据的关键步骤,其质量直接影响专病数据库的可靠性和实用性。数据采集前应明确范围,确定采集回顾性历史数据、前瞻性数据还是二者结合。在肺结节/肺癌专病数据库中,根据“数据池”字段类型,数据采集主要分为两种方式:针对院内数据,利用 ETL 技术,从 HIS、检验信息系统(laboratory information system, LIS)等数据源自动提取相关数据存入数据库;针对院外数据,由专病管理师随访患者时采集相关数据,同时收集量表数据。此外,数据采集要注重患者隐私保护和伦理审查。相关工作严格遵守《个人信息保护法》

《数据安全法》等法律法规,所有采集方案经伦理委员会审核批准,确保符合医学伦理要求。

3.2.3 “数据池”构建 “数据池”构建是将采集的多源异构数据整合存储,形成结构化数据集合,为后续数据查询、分析和应用提供基础。由信息中心牵头第三方公司设计开发专病队列数据库系统,存储采集的院内和院外数据。因采集数据来自不同数据源,可能存在数据格式不一致、重复、缺失等问题,须对数据进行二次结构化、集成及清洗。全程管理中心专病数据库管理师负责此项工作,保证数据的规范性、完整性,为后续应用奠定基础。

3.2.4 数据治理 数据治理是专病数据库数据质量、安全性和合规性的重要保障,贯穿数据库建设和使用全过程。其涵盖数据标准管理、数据质量管理、数据安全等方面。临床团队和全程管理中心共同制定、完善统一数据标准,包括数据字典、数据编码规则等,确保数据一致性,并定期更新以适应医学知识发展。在数据质量管理方面,临床团队指派数据审核员从准确性、完整性、一致性、及时性等维度评估数据,定期检查数据库数据质量,并督导整改。为保证数据安全,采取多种措施:对敏感数据加密存储防泄漏;通过内网登录、用户认证、角色授权控制数据访问权限;定期由第三方公司备份数据库,应对数据丢失或系统故障。

4 专病全程管理数据库建设经验

4.1 基本情况

在医疗信息化与精准医疗时代背景下,肺结节/肺癌专病全程管理数据库采用三方共建共享模式,以疾病特点为导向,构建包含肺结节全程管理、肺部穿刺活检、肺部手术治疗、肺癌非手术治疗 4 个子库的全程管理数据库。该数据库打破传统数据管理局限,通过子数据库互联互通与数据协同,实现患者从发现肺结节到肺癌的治疗、康复等信息的系统追踪。

四川大学华西医院肺结节/肺癌专病全程管理数据库于 2022 年开始建立肺结节全程管理子数据库,2024 年 3 月 18 日正式新增肺部穿刺活检、肺部手术治疗及肺癌非手术治疗 3 个子数据库,形成

完整的肺结节/肺癌全程管理数据库。截至 2025 年 7 月 1 日共纳入患者 36 262 例，其基础数据支撑产生多篇科研文章，支持多项国家级和省部级课题。

4.2 职责清晰的三方团队

传统专病数据库建设多以临床科研需求为核心，采用临床兼职录入与技术被动开发的松散架构。以临床为主导的模式重诊疗轻数据规范，导致数据描述随意性大，且技术部门与临床缺乏专职协调机制，数据采集与维护难以持续，易形成“烂尾库”。三方共建共享模式通过精准分工、紧密配合，形成“标准制定 - 技术支撑 - 运营保障”完整链条。基于该模式的肺结节/肺癌专病数据库团队，由三方指派人员承担项目管理员、随访管理员、数据录入员、数据审查员、数据管理员和系统工程师角色。临床团队指派人员担任项目管理员、数据录入人员和数据审查员，负责提供标准化数据采集需求、录入并审核数据；全程管理中心指派人员担任随访管理员和数据管理员，负责随访追踪患者结局事件等信息、对接临床需求并推动其向技术实现快速转化；信息中心指派人员担任系统工程师，负责数据库编程、测试及系统功能开发。此模式打破传统部门壁垒，使临床、技术、管理形成合力，为数据库长期稳定运行提供组织保障。

4.3 分阶段联动的数据链（图 2）

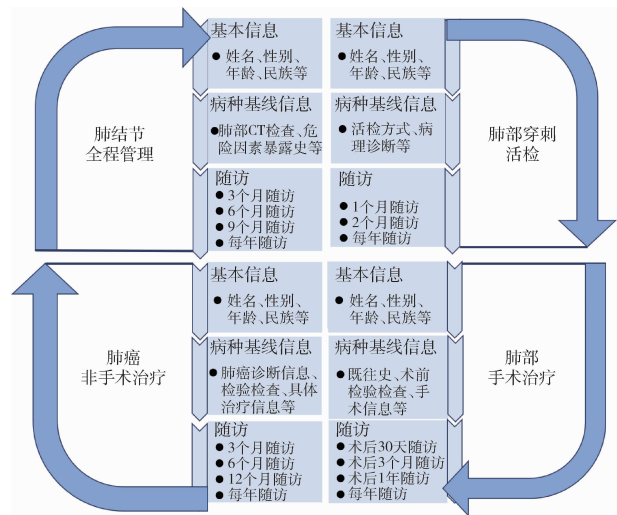


图 2 肺结节/肺癌专病数据库框架

传统专病数据库常采用单一库表，将筛查、诊疗、康复阶段割裂，导致数据碎片化、跨阶段关联困难，无法形成完整诊疗链条。该数据库基于全病程管理需求，打通院内多个系统，通过分阶段子数据库设计，院内院外多形式采集数据，实现数据采集与临床实践的适配。针对肺结节全程管理、肺部穿刺活检、肺部手术治疗及肺癌非手术治疗 4 个阶段的特点，制定数据录入和随访流程，确定患者从多通道入组的随访流程以及在每个阶段需要录入的数据。

数据库字段设计遵循“基础通用 + 特色专用”原则，为数据互联互通与精准应用奠定基础。基础字段如患者姓名、性别、家族史等基本信息，在 4 个子数据库中使用统一标准，确保数据在不同子库间流转兼容。特色字段深度贴合各子数据库特点：肺结节全程管理子数据库围绕结节发展全周期，设首次发现结节时间、影像学特征描述、历次随访结节大小变化等字段；肺部穿刺活检子数据库聚焦活检关键环节，含穿刺部位、穿刺方式、病理类型等字段；肺部手术治疗子数据库完整记录手术过程，有手术方式、手术时长、淋巴结清扫情况、术后病理等字段；肺癌非手术治疗子数据库精准追踪治疗效果，包含化疗、免疫等多种内科治疗药物信息、治疗评估等字段。

患者数据入组采用“主动纳入”与“被动筛选”双轨并行模式。医生可通过 HIS 手动选择不同诊疗阶段的患者入组，也可设置自动入组规则，系统根据患者病历字段自动判定，将符合各子数据库条件的患者纳入。患者病情变化时，可将其从初始入组子库转入新子库。此模式确保数据连续性，降低信息冗余，提升数据检索与分析效率，形成完整数据链，为全病程研究、诊疗方案优化提供真实标准数据源。

5 结语

四川大学华西医院构建三方共建共享的肺结节/肺癌全程管理数据库，突破了传统科研项目驱动的阶段性建库模式，破除了传统专病数据库建设壁垒，形成了可持续发展的新范式。该模式由临床团队、全程管理中心及信息中心协同发力，突破单一

主体建设局限,通过多维度整合临床资源、科研需求与技术优势,形成共建共享生态闭环。采用前瞻性与回顾性数据并行采集策略,打通院内诊疗与院外随访数据链路,实现患者全生命周期健康信息的无缝衔接,为后续数据应用筑牢根基。创新提出子数据库联动体系,将肺结节全程管理、肺部穿刺活检、肺部手术治疗、肺癌非手术治疗 4 大关键环节有机串联。各子库既有通用字段保障数据畅通,又通过个性化字段聚焦各阶段需求。患者数据可通过多元通道接入数据库,形成完整疾病发展轨迹图谱。这种全病程管理模式,既为临床医生提供全景式诊疗视角,助力其制定精准治疗方案;也为医学科研搭建起标准化数据平台,推动研究从单维分析迈向多维探索,有效解决传统数据分散、应用割裂难题。未来,该数据库将在现有创新成果基础上持续深耕,推动应用场景与功能迭代升级。一方面,开发患者端系统,通过四川大学华西医院官方微信公众号与 App 推送量表,便于患者在移动端自主填写评估状态量表,数据直接回传至数据库;另一方面,与院内人工智能 (artificial intelligence, AI) 全院统一随访平台打通接口,通过 AI 电话随访采集患者院外信息,并同步回传至数据库。肺结节/肺癌专病全程管理数据库的建设模式与实践经验将为我国医疗大数据领域发展提供有益借鉴,助力提升重大疾病防治能力,推动精准医学发展,为全国专病数据库建设规范化、协同化发展注入新动能。

作者贡献: 罗艺婷负责研究设计、论文撰写;杨先碧、杜春霖负责研究设计指导;张红、柴琪参与数据库构建与管理;邓懿负责论文审核;周永召负责提供指导、论文审核。

利益声明: 所有作者均声明不存在利益冲突。

参考文献

- 1 董方杰,胡建平,吴士勇.我国卫生健康信息互联互通 2.0 技术特征研究 [J]. 中国卫生信息管理杂志, 2023, 20 (1): 1-6.
- 2 WU Y J, HE S Y, CAO M D, et al. Comparative analysis of cancer statistics in China and the United States in 2024 [J]. Chinese medical journal, 2024, 137 (24): 3093-3100.
- 3 HANKEY B F, RIES L A, EDWARDS B K. The surveillance, epidemiology, and end results program: a national

resource [J]. Cancer epidemiology biomarkers & prevention, 1999, 8 (12): 1117-1121.

- 4 WINGO P A, RIES L A, GIOVINO G A, et al. Annual report to the nation on the status of cancer, 1973-1996, with a special section on lung cancer and tobacco smoking [J]. Journal of the national cancer institute, 1999, 19 (8): 675-690.
- 5 QIE S, ZHANG X, LUAN J, et al. Model development and validation for predicting small-cell lung cancer bone metastasis utilizing diverse machine learning algorithms based on the SEER database [J]. Medicine, 2025, 104 (12): e41987.
- 6 SETO Y, KAKEJI Y, MIYATA H, et al. National clinical database (NCD) in Japan for gastroenterological surgery: brief introduction [J]. Annals of gastroenterological surgery, 2017, 1 (2): 80.
- 7 BABA H. National clinical database (NCD) in Japan: clinical and social significance [J]. Annals of gastroenterological surgery, 2019, 3 (5): 462.
- 8 CONROY M C, LACEY B, BEŠEVIĆ J, et al. UK biobank: a globally important resource for cancer research [J]. British journal of cancer, 2023, 128 (4): 519-527.
- 9 DENG Y T, YOU J, HE Y, et al. Atlas of the plasma proteome in health and disease in 53026 adults [J]. Cell, 2025, 188 (1): 253-271.
- 10 白宇鸽,徐玲,段学宁,等.中国乳腺癌专病队列研究:临床队列的设计与初步结果 [J]. 中华流行病学杂志, 2020, 41 (12): 2046-2052.
- 11 王胤奎,李子禹,陕飞,等.我国早期胃癌的诊治现状——来自中国胃肠肿瘤外科联盟数据的启示 [J]. 中华胃肠外科杂志, 2018, 21 (2): 168-174.
- 12 WU Z, CHENG H, SHAN F, et al. In-hospital mortality risk model of gastric cancer surgery: analysis of a nationwide institutional-level database with 94277 Chinese patients [J]. Frontiers in oncology, 2019, 9 (10): 846.
- 13 姚宏伟,李心翔,崔龙,等.中国结直肠癌手术病例登记数据库 2019 年度报告:一项全国性登记研究 [J]. 中国实用外科杂志, 2020, 40 (1): 106-110, 116.
- 14 汪晓东,李希,何欣林,等.数据库研究第一部分:区域性医疗中心的结直肠癌与人群特征 [J]. 中国普外基础与临床杂志, 2019, 26 (2): 212-220.
- 15 李珂璇,孙振,邱辉忠,等.北京协和医院结直肠癌专病数据库的建立:单中心数据库经验 [J]. 协和医学杂志, 2023, 14 (3): 566-574.
- 16 阿依达娜·毛兰,张红,姜又琳,等.基于医疗大数据肺结节中医专病数据库建设的实践探索 [J]. 中国数字医学, 2025, 20 (3): 96-104.
- 17 吴晓龙,李浙民,陕飞,等.胃癌数据库建立和管理的质量化控制 [J]. 中华胃肠外科杂志, 2024, 27 (2): 132-136.
- 18 窦一峰,单淑香.基于 ETL 的心肺血管疾病专病数据库建设与实践 [J]. 医学信息, 2025, 38 (6): 69-73.
- 19 杨先碧,周永召,陈凤,等.基于“互联网+”的医护管一体化肺结节/肺癌全程管理模式构建及实践 [J]. 华西医学, 2024, 39 (4): 613-618.