

儿童注意缺陷多动障碍专病库构建与应用*

陆昕玥^{1,2,3} 黄 坚^{1,2,3} 陈凌栋^{1,2,3} 朱 珠^{1,2,3} 杨荣旺⁴ 李鑫鑫^{1,2,3} 赵永根^{1,2,3}
俞 刚^{1,2,3}

¹ 浙江大学医学院附属儿童医院数据信息部 杭州 310052

² 浙江-芬兰儿童健康人工智能联合实验室 杭州 310052

³ 国家儿童健康与疾病临床医学研究中心儿童健康医工信交叉研究中心 杭州 310052

⁴ 浙江大学医学院附属儿童医院心理科 杭州 310052

〔摘要〕 目的/意义 建立儿童注意缺陷多动障碍 (attention deficit hyperactivity disorder, ADHD) 专病库, 解决诊疗数据分散、数据质量偏低造成的数据使用效率低的问题。方法/过程 基于 2020 年 3 月—2024 年 4 月儿童 ADHD 临床诊疗数据, 利用自然语言处理技术实现数据后结构化, 根据行业标准、诊治指南实现专病库数据标准化。结果/结论 建立以 75 835 例 ADHD 患儿和 28 万人次就诊数据为基础的儿童 ADHD 专病库, 为 ADHD 人群基本特征分析、预测模型构建、优化资源配置、临床诊疗路径优化等应用方向提供数据支撑。

〔关键词〕 专病库; 儿童注意缺陷多动障碍; 数据治理

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2024.09.012

Construction and Application of Pediatric Attention Deficit Hyperactivity Disorder Sepcialized Disease Database

LU Xinyue^{1,2,3}, HUANG Jian^{1,2,3}, CHEN Lingdong^{1,2,3}, ZHU Zhu^{1,2,3}, YANG Rongwang⁴, LI Xinxin^{1,2,3}, ZHAO Yonggen^{1,2,3}, YU Gang^{1,2,3}

¹ Department of Data and Information, Children's Hospital, Zhejiang University School of Medicine, Hangzhou 310052, China; ² Sino-Finland Joint AI Laboratory for Child Health of Zhejiang Province, Hangzhou 310052, China; ³ Pediatric Medicine Engineering and Information Research Center, National Clinical Research Center for Child Health, Hangzhou 310052, China; ⁴ Department of Psychology, Children's Hospital, Zhejiang University School of Medicine, Hangzhou 310052, China

〔Abstract〕 **Purpose/Significance** The establishment of a specialized database for children with attention deficit hyperactivity disorder (ADHD) aims to address the issues of low data usage efficiency caused by scattered and low-quality diagnostic data. **Method/Process** Based on clinical diagnostic data of children with ADHD from March 2020 to April 2024, natural language processing (NLP) techniques are utilized to realize the post-structuring of the data. Furthermore, standardization of the database is performed according to industry standards and treatment guidelines. **Result/Conclusion** The specialized database for children with ADHD is established, comprising data from 75 835 ADHD children and 280 000 clinical visits. This database supports various applications such as the analysis of basic characteristics of the ADHD population, prediction model construction, resource allocation optimization, and clinical pathway optimization.

〔Keywords〕 specialized disease database; attention deficit hyperactivity disorder (ADHD) in children; data governance

〔修回日期〕 2024-07-29

〔作者简介〕 陆昕玥, 硕士, 助理工程师; 通信作者: 俞刚, 教授级高级工程师。

〔基金项目〕 国家重点研发计划项目 (项目编号: 2023YFC2706400)。

1 引言

注意缺陷多动障碍（attention deficit hyperactivity disorder, ADHD）多发于学龄儿童，是一种神经发育障碍性疾病，以注意力不集中、多动和冲动行为为主要临床特征，严重影响儿童的社会交往能力和学习能力^[1-2]。据统计，我国 ADHD 儿童患病率约 6.26%，约有 2 300 万儿童及其家长正面临该疾病的困扰^[3]，且 70% 的儿童功能障碍会持续至成年^[4]。该疾病对儿童身心健康影响严重，疾病持续时间和治疗周期长，给社会和患儿家庭造成巨大的经济负担^[5]。2021 年 10 月国家卫生健康委员会提出，要加强儿童精神心理专科建设，积极干预儿童心理行为发育异常^[6]。利用大数据进行各类专病研究被认为是深入探索专病特点、实现个性化诊疗的高效手段^[7]。专病库是专科病种数据管理和知识管理的有效手段^[8]。然而目前国内大部分专病库建设聚焦于器质性疾病，针对儿童功能性疾病尤其是 ADHD 的专病库较少。本研究依托浙江大学医学院附属儿童医院（以下简称医院）临床数据中心，建立儿童 ADHD 专病库，进而辅助临床医生实现 ADHD 儿童特征、影响因素和治疗效果分析，为 ADHD 儿童慢病管理和治疗路径优化提供有力的数据支撑。本研究遵守《赫尔辛基宣言》，经浙江大学医学院附属儿童医院伦理委员会批准（〔伦〕审编号：2023-IRB-0287-P-01）。

2 现状分析

2.1 信息共享与数据集成的局限性

虽然医院已经通过构建基于企业服务总线的集成平台和临床数据中心（clinical data repository, CDR），初步实现了各业务系统之间的信息共享与互联互通，满足了医生对临床数据的查阅需求，但

是当前的 CDR 在数据管理方面仍存在局限性。各系统逻辑组织方式与数据格式不同，而 CDR 仅对院内业务系统数据进行基础集成和展示，缺乏有效的数据治理机制，导致各类疾病相关数据中存在大量重复值和异常值，数据结构不一，数据质量参差不齐^[9]。这些问题严重制约数据的有效利用，难以支撑临床诊疗路径的优化和科研工作的深入开展。

2.2 ADHD 患儿诊疗数据的复杂性

与器质性疾病患儿相比，ADHD 患儿的诊疗过程具有特殊性和复杂性。除了常规的医学检验和检查项目外，医生还需要儿童、家长、老师等多方填写多种评估量表，如皮博迪图片词汇测验、注意力缺陷多动障碍评定量表（Swanson, Nolan and Pelham-IV rating scales, SNAP-IV）、康奈氏儿童行为量表（Conners parent symptom questionnaire, PSQ）等。医生通过量表评估结果，量化和比较儿童的心理状况，辅助疾病诊断。然而，目前医院中量表数据的构成和管理面临诸多挑战。量表数据源自多个院内外系统，结构复杂；早期量表数据多以纸质形式存在，医生需手动将其录入电子病历系统，大量数据以非结构化形式存储，难以直接用于临床科研。此外，临床观察和访谈作为评估 ADHD 患儿的重要步骤^[10]，其记录同样以非结构化数据的形式存储在电子病历中。

3 儿童 ADHD 专病库构建流程

儿童 ADHD 专病库旨在提供一个高效、安全、可扩展的数据存储和管理平台，以支持 ADHD 患儿的全病程管理和临床研究。儿童 ADHD 专病库分为数据源层、CDR 数据支撑层、数据治理层与应用层。数据安全管理体系则贯穿整个专病库构建流程，包括权限管理、去隐私、安全审计等。儿童 ADHD 专病库技术架构，见图 1。

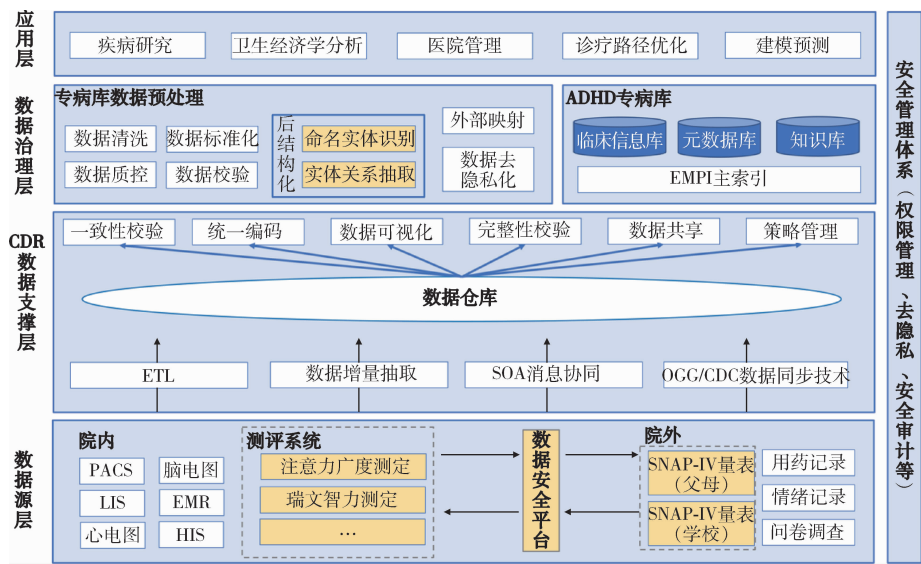


图1 儿童ADHD专病库技术架构

3.1 数据源分析

儿童ADHD专病库中的数据来源于医院各业务系统。院内数据主要包括影像存储与传输系统、实验室信息管理系统、电子病历系统等系统中的患者诊疗数据。院外数据主要包括随访系统数据。此外，由于儿童的特殊性，家长和学校教师需要根据儿童日常表现填写量表。因此，测评系统包括院内和院外数据，通过数据安全平台保障患者数据的安全传输。院内网中的量表由儿童在就诊时填写，而公网的量表主要由家长和学校老师在手机端填写。

3.2 CDR数据集成

遵循实时数据集成、数据标准化和多值溯源原则，采用全量抽取方式抽取业务数据，保持源数据完整性。作为院内大数据处理平台，临床数据中心通过变化数据捕获技术，实现对源数据表的实时提取，捕获新增、删除、修改等操作。此外，利用抽取、转换和加载（extraction - transformation - loading, ETL）工具进行实时数据集成。数据集成的主要目标是将业务系统数据加载到数据仓库，不影响医院各业务系统生产库性能，提高数据检索效率。然后，再对数据进行完整性校验、一致性校验等操作，实现数据集成和初步处理。

3.3 专病库构建

3.3.1 专病库设计 儿童ADHD专病库基于诊治指南制定患者纳入、排除标准，明确数据范围。儿童ADHD的主要诊断标准：在注意缺陷和（或）多动、冲动症状至少一个维度的9项描述中，出现至少6项符合描述的情况；注意缺陷和（或）多动、冲动症状至少存在于2个场合（如家庭、学校等）；注意缺陷和（或）多动、冲动症状持续出现至少6个月；有单个或多个功能损害（如学业、社会功能等）；排除其他精神障碍或神经系统疾病。专病库的数据纳入标准：年龄18周岁以下；临床访谈和心理社会评估完善；既往身体健康。专病库的数据排除标准：量表测评结果缺失；既往有神经系统器质性疾病或广泛发育障碍的患儿。专病库基于《卫生健康信息数据元标准化规则》（WS/T 303—2023）、《卫生健康信息数据元值域代码》（WS/T 364—2023）等国家卫生健康委员会行业数据标准设计，将业务数据库中的医疗词汇与外部标准词汇映射，为后续数据的高效检索奠定基础。儿童ADHD专病库根据诊疗规范和医院专科诊疗路径设计，主要由患者就诊信息、人口学信息、病历信息、测评量表、检查等8个一级模块构成。其中，测评量表模块涵盖了儿童ADHD诊断过程中常用的

量表,例如:注意力广度测定、Weiss 功能缺陷评定、康奈氏儿童行为量表等。儿童 ADHD 专病库部

分字段与数据源,见图 2。

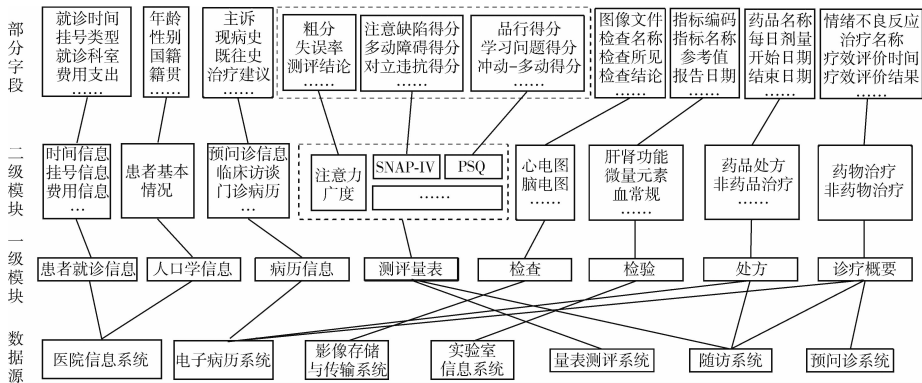


图2 儿童 ADHD 专病库部分字段与数据源分析

3.3.2 多源异构数据整合 儿童 ADHD 专病库数据来源于多个业务系统。其中,较早的心理测评系统未接入网络环境,无法直接与患者其他数据关联。此外,部分业务系统以门诊号作为患者标识,造成有多个门诊号的患者数据无法整合。因此,需对业务数据基于院内患者的统一编码,完善患者主索引。首先,对较早的心理测评系统数据导出后台数据,通过门诊号初步与院内系统关联。然后,通过对身份证号、出生日期、地址等的规范化处理,完成患者标识的标准化处理。应用概率匹配等算法对不同业务系统的患者标识信息重新组织,形成患者唯一标识编码,即企业级患者主索引识别码,合并同一患者就诊数据,实现多源异构数据的融合。

3.3.3 文本数据后结构化 非结构化数据主要来源于电子病历系统中的门诊病历、诊疗补充记录,影像存储与传输系统中的脑电图数据、检查报告等。此外,较早的儿童心理测评结果数据由接诊医生以文本形式记录在电子病历中。由于非结构化数据记录形式灵活多变,直接与儿童 ADHD 相关标准医学术语映射较为困难。因此,在专病库数据生产过程中,运用双向长短时记忆网络、条件随机场等深度学习算法,对病历文书等非结构化数据进行中文分词、命名实体识别、实体关系抽取等操作,实现文本数据的后结构化。

3.3.4 数据治理 CDR 中的缺失项和重复值会影响专病库的数据分析结果和利用效率。因此,

在儿童 ADHD 专病库数据生产前,需要进行数据清洗、数据标准化和数据校验,以实现数据质量控制。(1) 数据清洗。专病库字段根据临床医生的需求设定。对于符合纳入、排除标准的异常数据,采用数据清洗方法提高准确性和一致性。主要方法包括时间格式清洗和浮点数据清洗等^[11]。(2) 数据标准化。由于专病库的时间跨度长、数据来源多,原始数据中包含大量非标准化的医疗术语。因此,在完成数据清洗后,基于国际相关数据标准、国家卫生健康委员会行业数据标准和 ADHD 诊治指南等外部标准,对诊断、症状、检验报告等患者数据标准化处理,与院内数据映射,形成统一编码,构建标准数据元。专病库参考标准,见表 1。依据《国际疾病分类第十一次修订本(ICD-11)》《医学系统命名法——临床术语(SNOMED CT)》,归纳整理诊断数据中出现频次最高的 50 个诊断的疾病编码和书写方式,形成儿童 ADHD 专病库的诊断数据字典。同时,对后结构化的儿童 ADHD 症状描述分类,提取与注意缺陷、多动和冲动以及其他共患病相关的关键信息。此外,检验数据存在同一指标名称不一致、参考范围变更、单位不统一等问题,也需要数据标准化处理,制定统一标准,形成检验数据字典。(3) 数据校验。为保证专病库数据的一致性,对数据进行基于规则或算法的验证,包括检查年龄和出生日期是否匹配等逻辑校验。

表1 儿童 ADHD 专病库参考标准

标准类型	名称
国际相关数据标准	《国际疾病分类第十一次修订本（ICD-11）》
	《医学系统命名法——临床术语（SNOMED CT）》
国家卫生健康委员会 行业数据标准	《卫生健康信息数据元标准化规则（WS/T 303—2023）》
	《卫生健康信息数据元值域代码（WS/T 364—2023）》
	《常用临床医学名词（2023年版）》
儿童 ADHD 诊治指南	《精神障碍诊断与统计手册（第5版）（DSM-V）》
	《注意缺陷多动障碍早期识别、规范诊断和治疗的儿科专家共识》

3.4 数据安全管理体系

由于公众对 ADHD 的认知程度较低，且污名现象非常普遍^[12-13]，部分 ADHD 患儿家长担忧儿童隐私泄漏，不愿配合部分量表的填写，从而影响 ADHD 患儿的诊断和预后。为保障数据安全性，专病库构建从数据传输、应用等方面建立完善的安全管理体系。首先，通过公网访问的系统已全部实现服务器本地部署，将医疗数据的存储和操作限制在院内网环境。数据安全平台对院内各系统接口加密和封装，并实时监控数据传输情况。患儿院内 ID、

门诊号、家长手机号等信息属于可追溯的潜在患者标识。针对部分需要在公网环境中由家长和老师填写的测评量表，数据安全平台在接口封装时，对儿童身份标识脱敏处理，降低院外隐私数据泄漏风险。其次，在儿童 ADHD 专病库数据抽取前，对儿童隐私信息的字段采用混合屏蔽、数据偏移、数字随机化等方式实现信息脱敏处理，进一步提升数据安全性^[14]。最后，专病库的访问实行授权登录和权限分级管理模式。院内医生账号仅支持数据的检索和查阅。专病库数据的导出须经过办公审批流程，保障数据应用层的安全。

4 儿童 ADHD 专病库数据应用

4.1 人群基本特征分析

目前，专病库共收集 ADHD 儿童就诊数据超 28 万条。所有儿童均符合《精神障碍诊断与统计手册（第5版）》^[15] 诊断标准，共 75 835 例，其中男童 59 113 例（约占 77.9%），女童 16 722 例（约占 22.1%）。专病库中有 16 632 例儿童存在共患病（占 21.9%）。共病主要包括社交障碍、对立违抗性障碍、童年情绪障碍、孤独症谱系障碍、癫痫等。专病库中 ADHD 患儿初次确诊时的平均年龄为 8.12 岁（男童为 8.1 岁，女童为 8.17 岁）。根据专病库中的数据统计显示，约 80% 患者选择在心理科就诊。浙江省内儿童占就诊人数 78%，其中 42.3% 来自杭州市。ADHD 患儿初次确诊年龄分布与区域分布，见图 3。

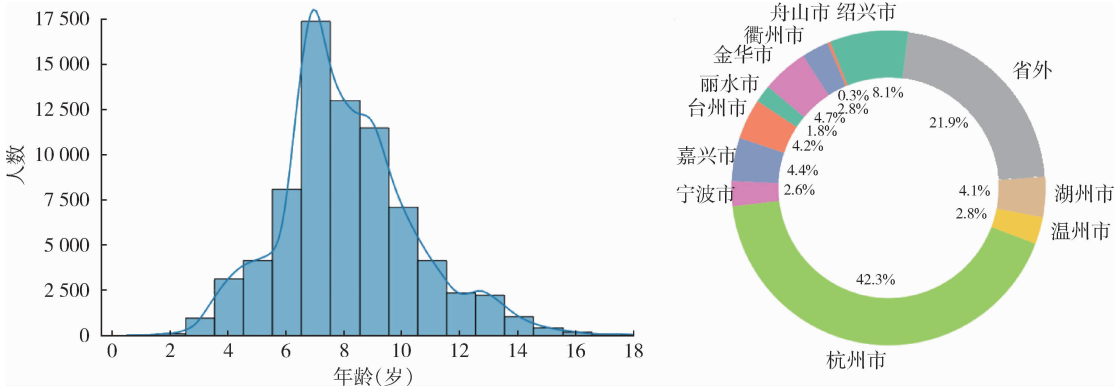


图3 ADHD 患儿初次确诊年龄与区域分布

4.2 疾病预测模型构建

专病库中，初次确诊年龄为6—12岁的学龄期儿童占总病例数的82.26%，6岁以下学龄前儿童占总病例数的13.96%，卡方检验结果显示学龄期儿童病例明显多于学龄前儿童($\chi^2 = 21\,172$, $P < 0.05$)。专病库显示ADHD患儿症状主要表现在学校中，如容易分心、打断教师讲课、无法按时完成作业。相比于学龄前ADHD患儿，学龄期ADHD患儿的症状更易被老师识别。然而，ADHD患儿在学龄前就会表现出部分症状特征，且学龄前儿童的神经发育可塑性更强，治疗效果更佳^[16]。专病库数据可为ADHD的早期预防、早期干预、早期治疗提供科学依据。例如，临床医生可以通过专病库数据，对学龄期儿童开展回顾性研究，提取患者基本信息、检验、影像学信息等数据，运用大数据分析和机器学习技术，探讨患儿个体和家庭影响因素，预

测学龄前儿童患病风险。

4.3 优化资源配置，改善患者体验

自2020年以来，医院诊断为ADHD的儿童人数逐年增加。2021年平均每月就诊人数为5 494人次；2022年增至5 516人次；2023年达到6 778人次，较2021年增加23.4%。2024年4月的ADHD患儿就诊量更是达到8 302人次，创月就诊人次数的新高。每月ADHD患儿就诊情况，见图4。这一趋势可能是由于学校和家长对儿童ADHD的重视程度逐渐提高，以及近年来儿童ADHD患病率的上升^[17]。通过分析ADHD患儿家长的就诊行为，医院可以优化资源配置，识别潜在问题，提高医疗服务质量和患者体验。例如，帮助医院预测就诊高峰、合理安排医生数量和门诊时间、提高诊疗效率。这些措施将有助于提升医院的整体服务水平，增强患者的满意度和信任度。



图4 就诊日期与就诊人次热力图

4.4 临床诊疗路径优化

专病库中ADHD患儿平均持续治疗时间为57.4天，平均就诊次数为2.83次。患者平均单次诊疗支出（包括医保统筹基金支付金额和患者自费金额）约为520.3元。初诊时，患者主要支出集中于量表测评、检验项目和检查项目，复诊支出则主要来自药品费用。《注意缺陷多动障碍早期识别、规范诊断和治疗的儿科专家共识》指出，ADHD患儿的治疗周期通常较长，包括药物治疗和行为干预，

需要家庭、学校等多方配合，制定长期治疗计划，按照慢性病管理策略进行管理^[3]。但医院患者的持续复诊时间较短且就诊次数少，可能存在治疗依从性不足问题。因此，医院管理人员可进一步研究患者治疗依从性不足的原因，如经济因素、家长认知和社会支持等，完善治疗策略。通过诊后宣教、随访、增加互联网医院线上问诊等方法，提升ADHD患儿诊疗质量。同时加强医生与家长的沟通和家长培训，提高治疗依从性，改善治疗效果和患者预后。

5 结语

本研究基于医院临床数据中心, 构建儿童 ADHD 专病库, 实现了 ADHD 患儿全病程数据的规范化收集、存储和管理, 为儿童 ADHD 的诊断和治疗提供了可靠的数据支撑, 也为 ADHD 疾病的临床研究、医院管理和卫生经济学研究提供了重要基础。临床研究方面, 儿童 ADHD 专病库的建立使研究人员能够获取全面、详尽的数据, 开展更深入的病例研究。此外, 通过分析患者的诊断、量表、影像等多模态数据, 借助人工智能技术, 可以建立预测模型, 分析 ADHD 的致病原因和药物治疗效果。医院管理方面, 儿童 ADHD 专病库提供了患者初诊、复诊等全面的就诊数据, 医院可以据此改进医生排班、线上咨询等患者管理措施, 优化诊疗流程, 提高医疗服务质量, 为患者提供更好的就医体验。卫生经济学研究方面, 儿童 ADHD 专病库的数据为评估 ADHD 患儿的社会经济负担和医疗成本提供了重要依据, 从而为卫生政策的制定提供数据支撑。

依托医院注意缺陷与多动障碍诊疗中心, 以儿童 ADHD 专病库为基础, 逐步纳入更详细的 ADHD 共病诊疗数据。首先, 专病库建设工作将致力于提升数据质量, 确保专病库数据的可靠性和一致性。其次, 专病库将逐步扩大数据来源, 通过加强与其他医疗机构的合作, 丰富内容。最后, 将进一步强化数据隐私和安全保护, 确保患者数据的保密性和合法性。随着人工智能和数据挖掘技术的发展, 专病库将在未来发挥更大的作用, 为个性化治疗和精准医疗提供强有力的数据支持, 从而为患者提供高质量的医疗服务。

作者贡献: 陆昕玥负责论文构思、数据收集与分析、论文撰写与修订; 黄坚负责统计学处理、结果分析、文章修订; 陈凌栋、朱珠、杨荣旺、李鑫鑫、赵永根负责研究可行性分析、数据收集、数据库建设; 俞刚负责论文审核。

利益声明: 所有作者均声明不存在利益冲突。

参考文献

- 1 POSNER J, POLANCZYK G V, SONUGA - BARKE E. Attention - deficit hyperactivity disorder [J]. *Lancet*, 2020, 395 (10222): 450 - 462.
- 2 江文庆, 杜亚松. 注意缺陷多动障碍与认知迟缓特质 [J]. *中国儿童保健杂志*, 2021, 29 (6): 632 - 635.
- 3 中华医学会儿科学分会发育行为学组. 注意缺陷多动障碍早期识别、规范诊断和治疗的儿科专家共识 [J]. *中华儿科杂志*, 2020 (3): 188 - 193.
- 4 RIGLIN L, COLLISHAW S, THAPAR A K, et al. Association of genetic risk variants with attention - deficit/hyperactivity disorder trajectories in the general population [J]. *JAMA psychiatry*, 2016, 73 (12): 1285 - 1292.
- 5 张玲, 刘寰忠. 儿童注意缺陷多动障碍遗传学研究进展 [J]. *中国实用儿科杂志*, 2023, 38 (8): 580 - 584.
- 6 国家卫生健康委员会. 关于印发健康儿童行动提升计划 (2021—2025 年) 的通知 [EB/OL]. [2024 - 05 - 22]. https://www.gov.cn/zhengce/zhengceku/2021 - 11/05/content_5649019.htm.
- 7 KRUSE C S, GOSWAMY R, RAVAL Y, et al. Challenges and opportunities of big data in health care: a systematic review [J]. *JMIR medical informatics*, 2016, 4 (4): e38.
- 8 靳淑雁, 王爽, 黄琼, 等. 基于乳腺癌专病库的知识图谱构建研究 [J]. *医学信息学杂志*, 2023, 44 (12): 65 - 70.
- 9 侯丽, 洪娜, 李露琪, 等. OHDSI 通用数据模型及医学术语标准国内应用现状分析 [J]. *医学信息学杂志*, 2020, 41 (2): 2 - 10.
- 10 冯廷勇, 王雪珂. 注意缺陷多动障碍儿童多模态评估的现状与建议 [J]. *中国儿童保健杂志*, 2024, 32 (6): 586 - 590, 601.
- 11 李鹏, 聂刚, 刘庆金, 等. 基于医院科研大数据中心的专病数据库建设实践 [J]. *中国数字医学*, 2023, 18 (9): 85 - 89, 120.
- 12 李成哲, 石宇婧, 宗亚辉, 等. 基于连续体信念的精神障碍污名化干预探索 [J]. *心理科学进展*, 2023, 31 (2): 288 - 300.
- 13 陈碧涵. 家庭积极行为支持在 ADHD 儿童问题行为干预中的应用研究 [D]. 广州: 广州大学, 2023.
- 14 唐凯, 张国明, 楚胜翔. 基于数据脱敏技术的大数据隐私安全应用与实践 [J]. *中国卫生信息管理杂志*, 2022, 19 (3): 436 - 442.
- 15 American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed.) [EB/OL]. [2024 - 05 - 22]. <https://archive.org/details/APA - DSM - 5>.
- 16 徐磊, 刘芳, 黄翔, 等. 学龄前儿童注意缺陷多动障碍早期特征的研究进展 [J]. *妇儿健康导刊*, 2024, 3 (4): 11 - 14, 20.
- 17 涂梦璐. 学龄前 ADHD 儿童执行功能特征及规则游戏干预的有效性研究 [D]. 上海: 华东师范大学, 2024.