

· 标准与规范 ·

医学检验标准数据集要素专家共识 (2026 版)

医疗辅助技术(医学检验)国家临床医学研究中心 中国医师协会检验医师分会智能
检验学组 医学检验人工智能创新协同工作组

通信作者:尚红,中国医科大学附属第一医院检验科 医疗辅助技术(医学检验)国家
临床医学研究中心,沈阳 110001,Email:hongshang100@hotmail.com;陈瑜,浙江大学
医学院附属第一医院检验科,杭州 310003,Email:chenyuzy@zju.edu.cn

【摘要】 随着人工智能技术的迅猛发展,尤其是以大型模型为代表的生成式人工智能技术的兴起,医学检验领域正逐步迈入智能化发展的新阶段。在此背景下,医学检验标准数据集作为大模型训练的重要基础、模型性能评价的关键基准以及应用场景开发的重要支撑,其作用愈发凸显。为加强相关专业人员对医学检验标准数据集相关概念与构建规范的共识认知,并回应我国医学检验多中心数据整合难、信息系统异构及数据质量差异显著等现实问题,在系统借鉴国内外相关共识和指南的基础上,医疗辅助技术(医学检验)国家临床医学研究中心、中国医师协会检验医师分会智能检验学组以及医学检验人工智能创新协同工作组组织多领域专家,围绕医学检验标准数据集的概念、构建意义与目标、构建原则、构建方法、伦理与安全,及其价值与应用五个方面制订了本专家共识,提出具有本土适配性的医学检验标准数据集构建与治理框架,以期推动我国医学检验智能化与规范化发展。

【关键词】 医学检验; 人工智能; 数据集; 专家共识; 全生命周期; 质量控制; 伦理治理

实践指南注册:国际实践指南注册与透明化平台(PREPARE-2025CN1579)

Expert consensus on the essential elements for standardized medical laboratory datasets (2026 edition)

National Clinical Research Center for Medical Auxiliary Technology (Laboratory Medicine); Intelligent Laboratory Medicine Group of the Laboratory Physicians Branch of the Chinese Medical Doctor Association; Collaborative Innovation Working Group for Artificial Intelligence in Laboratory Medicine
Corresponding authors: Shang Hong, National Clinical Research Center for Medical Auxiliary Technology (Laboratory Medicine), Department of Laboratory Medicine, the First Hospital of China Medical University, Shenyang 110001, China, Email: hongshang100@hotmail.com; Chen Yu, Department of Laboratory Medicine, the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, China, Email: chenyzy@zju.edu.cn

【Abstract】 With the rapid advancement of artificial intelligence (AI) technologies, particularly the emergence of generative AI represented by large-scale models, the field of laboratory medicine is gradually entering a new stage of intelligent development. In this context, standardized laboratory medicine datasets have become increasingly critical as a fundamental basis for large model training, a key benchmark for model performance evaluation, and an essential support for the development of application scenarios. To strengthen consensus among professionals

DOI: 10.3760/cma.j.cn112137-20260115-00174

收稿日期 2026-01-15 本文编辑 张志巍

引用本文:医疗辅助技术(医学检验)国家临床医学研究中心,中国医师协会检验医师分会智能检验学组,医学检验人工智能创新协同工作组.医学检验标准数据集要素专家共识(2026版)[J].中华医学杂志, 2026, 106(21): 2107-2127. DOI: 10.3760/cma.j.cn112137-20260115-00174.



中华医学杂志社
Chinese Medical Association Publishing House

版权所有 违者必究



regarding the related concepts and construction specifications of standardized laboratory medicine datasets, and to address practical challenges in China such as difficulties in multicenter data integration, heterogeneous information systems, and substantial variability in data quality, National Clinical Research Center for Medical Auxiliary Technology (Laboratory Medicine), together with the Intelligent Laboratory Medicine Group of the Laboratory Physicians Branch of the Chinese Medical Doctor Association and the Collaborative Innovation Working Group for Artificial Intelligence in Laboratory Medicine, organized experts from multiple disciplines. Drawing systematically on relevant domestic and international guidelines and consensus documents, this expert consensus was developed across five key domains: the concept of standardized laboratory medicine datasets, construction significance and objectives, construction principles, construction methodologies, ethics and security, as well as their value and applications. This consensus proposes a laboratory medicine dataset construction and governance framework tailored to the Chinese context, with the aim of promoting the intelligent and standardized development of laboratory medicine in China.

【Key words】 Medical laboratory science; Artificial intelligence; Dataset; Expert consensus; Full life cycle; Quality control; Ethical governance

Practice guideline registration: Practice Guideline Registration for Transparency (PREPARE-2025CN1579)

在医疗领域,医学检验扮演着数据生产者的角色,其产生的数据对临床决策、健康管理具有至关重要的作用,约占临床决策依据的70%^[1]。随着人工智能(artificial intelligence, AI)技术的迅猛发展,尤其是以大型模型为典型代表的生成式AI技术的兴起,医学检验领域已然步入智能化发展的新阶段。在这一背景下,医学检验标准数据集的重要性体现在三个层面:首先,作为构建大型模型的基础,医学检验标准数据集确保了医学检验生成式AI训练数据的高质量与可靠性;其次,作为模型评估的标准,为生成式AI技术的成果提供了可对比和可量化的评价方法;最后,它作为应用实施的基准,促进了医学检验核心场景中智能应用开发的标准化。鉴于在既往工作中遇到的问题,为解决相关专业人员对于医学检验标准数据集构建领域概念认识的一致性,并加快医学检验大数据及AI技术的发展,医疗辅助技术(医学检验)国家临床医学研究中心、中国医师协会检验医师分会智能检验学组及医学检验人工智能创新协同工作组整合当前国内外最新研究进展与临床证据,联合制订《医学检验标准数据集要素专家共识》(以下简称“本共识”),以期提高医学检验大数据及AI领域研究质量并促进未来多学科相互合作。

第一部分 共识制订方法

1. 共识牵头单位及注册:本共识由医疗辅助技术(医学检验)国家临床医学研究中心、中国医师协会检验医师分会智能检验学组及医学检验人工智

能创新协同工作组联合发起,启动时间为2025年6月,并在国际实践指南注册与透明化平台(<https://www.guidelines-registry.org>)注册,注册号:PREPARE-2025CN1579。

2. 共识目标人群与使用者:本共识应用的目标人群为医学检验及相关学科的临床、科研与管理人员,以及从事医学AI研发和应用的专业人士。主要使用者为各级医疗机构的检验实验室、科研机构、行业企业与教育培训单位,可为临床诊疗、科研创新、AI模型训练与评估、实验室质量管理及行业标准制订提供参考与依据。

3. 共识制订工作组:本共识制订小组包括共识领导组、共识制订专家组、执笔专家组,以及执笔统筹组。共识领导组的主要职责是:统筹安排整个共识撰写工作。共识制订专家组由医学检验、指南方法学、AI及大数据等相关领域专家组成。共识制订专家组的主要职责是:制订本共识的内容范围和文本结构,对推荐意见进行审核和讨论,并给出修改建议,负责对本共识进行整体把握。共识执笔专家组的主要职责是:按照制订专家组制订的文本结构,分别筛选相关文献和初步凝练推荐意见,并撰写共识全文。共识执笔统筹组成员为3人,负责本共识的注册、组织本共识撰写和讨论工作会议以及详细记录本共识制订的整个过程。

4. 证据检索和质量评价:系统检索国内外权威数据库,包括PubMed、Web of Science、万方数据知识服务平台和中国知网等,以及与医学检验大数据相关的政府、行业主管部门、学会/协会网站文件,检索时间为建库至2025年10月,检索不限定发表



语言。英文检索词主要为“laboratory medicine”“medical laboratory”“clinical laboratory”“laboratory test”“dataset”等,中文检索词包括“检验医学”“医学检验”“实验室医学”“数据集”等。依托预设的纳入与排除标准,由2名研究者独立筛选文献:优先纳入与医学检验数据集构建、数据标准化及AI应用相关的文献,研究类型包括相关指南、专家共识及系统评价等。排除研究内容无关、无法获取全文、发表语言非中文或英文以及重复发表的文献。

在系统文献检索基础上,对纳入研究进行方法学质量评价。对于系统综述与荟萃分析等文献,采用使用系统评价偏倚风险评价工具(A MeaSurement Tool to Assess systematic Reviews, AMSTAR)进行质量评价;对于观察性研究,采用纽卡斯尔-渥太华量表(Newcastle-Ottawa Scale, NOS)进行质量评价。鉴于本共识纳入文献主要包括指南、专家共识、方法学研究及技术规范性文件,传统临床研究证据相对有限,因此在质量评价基础上,结合文献来源、方法学规范性及专家经验进行综合判断。

5. 证据质量和推荐强度:本共识采用推荐意见分级评价、制订及评估(Grading of Recommendations, Assessment, Development and Evaluation, GRADE)方法对证据质量和推荐意见强度进行分级(表1)。

表1 推荐意见分级评价、制订及评估(GRADE)

证据质量与推荐强度分级	
分级	定义说明
证据质量分级	
A(高)	现有研究结果高度可信,进一步研究极有可能不会改变对效应大小的判断
B(中)	对研究结果具有中等信心,进一步研究可能对效应估计产生一定影响
C(低)	对研究结果的可信度有限,进一步研究很可能改变现有结论
D(极低)	研究证据非常不确定,观察结果与真实效应之间可能存在显著差异
推荐强度分级	
1(强推荐)	明确表明干预措施的预期获益大于潜在风险或不良影响
2(弱推荐)	干预措施的获益与风险关系尚不确定,或不同证据结果存在差异

6. 推荐意见的形成:本共识制订过程中,遵循多学科参与原则遴选专家,涵盖医学检验、AI与信

领域。专家遴选综合考虑专业背景、研究或实践经验及学术代表性,以保证推荐意见在临床可行性、技术合理性与伦理合规性等方面的平衡。专家组基于系统文献检索和相关实践经验,经过1轮专家意见征集,最初确定了标准数据集要素应涵盖的5大模块及5个子模块。共识撰写组成员基于指南、专家共识、系统评价及相关方法学研究等证据,分成不同领域,根据专家组预先确定的范围,初拟了共识意见和证据与解释。统筹组整合共识意见后,组织1轮会前专家组评审和修改意见汇总,以及2轮共识专家组及撰写组共同参与的线上讨论会议,对共识进行讨论与修改。其中,会前共收集专家组评审和修改意见132处,第1轮线上会议对16条推荐意见提出修改建议,第2轮线上会议对4条推荐意见提出修改意见,初步形成了31条推荐意见。修改完成后,统筹组通过问卷调查的形式,在中国医师协会检验医师分会智能检验学组中进行2轮德尔菲法问卷调查,调研人数41人,2轮问卷有效回收率分别为90.9%、87.8%。其中,调查问卷共分为三部分,分别为专家基本情况收集、基于李克特量表的推荐意见评分以及总体反馈等。若专家对某个推荐意见的评分低于5分,则需说明具体原因及修改建议。基于第一轮德尔菲问卷调查的评分意见,对本共识初稿中有专家异议的6条推荐意见进行修改情况说明并进行第二轮德尔菲调查,最终形成27条推荐意见。

德尔菲调查问卷内容主要包括针对每条推荐意见的李克特量表评分及可自由填写的意见和建议区域。针对每一条推荐意见,专家采用李克特量表评分,满分7分,7分表示非常同意,6分表示同意,5分表示一般同意,4分表示不确定,3分表示不太同意,2分表示不同意,1分表示完全不同意。本共识设定:针对单条推荐意见,评分 ≥ 6 分的专家超过70%,则该条推荐意见达成共识。专家推荐程度以“共识度”标注,共识度=评分 ≥ 6 分的专家/总参评专家人数 $\times 100\%$ 。

7. 利益冲突:所有作者均声明不存在利益冲突。

第二部分 医学检验标准数据集的概念、构建意义及目标

本章的核心目标为明确医学检验标准数据集的定义与高质量内涵,阐述其在数据治理、多中心



整合以及 AI 模型开发与评测中的基础性作用,并提出标准数据集的应用场景和覆盖范围。

推荐意见 1: 医学检验标准数据集是指围绕明确的临床、科研与 AI 任务,在一致的数据架构约束与编码规范下,将多源检验数据与关键元数据经标准化与质量控制形成的,具备完整性、规范性、准确性、均衡性与时效性的高质量可复用的数据集。(1A, 共识度 93.3%)

解读与证据: 近年来,我国将构建高质量数据集提升到战略高度:从《新一代人工智能发展规划》^[2]到“数据要素×三年行动计划(2024—2026)”^[3]、再到《数字中国建设 2025 年行动方案》^[4],陆续出台了一系列政策持续强调以数据驱动的创新体系,并明确提出在医疗等重点行业建设高质量数据集。伴随“以数据为中心”的范式转向,行业共识从“越多越好”转向“高质可用”,并将可评估、可复用、可验证作为数据集建设的核心要求。

医学检验数据的独特性在于多源异构与强情境依赖,即同名项目可因检测方法、标本类型、仪器与试剂批号、参考区间、人群结构而呈现显著差异。如果不在“数据架构约束与编码规范”层面固化表达,后续研究设计与模型训练便容易把流程性噪声误当作生物学信号。所谓的“标准数据集”,并非把原始记录“搬到”同一文件中,而是在一致的最小元数据集框架内,将检验项目与结果、单位与参考区间、标本与方法学、质控状态与时间戳等关键要素连同其生成条件一并沉淀,通过统一口径与规则化处理形成可直接用于研究与建模的“数据制品”。其中,元数据描述的是数据集的来源、结构、方法学及时间戳等。这一框架的数据质量并非由单一特性决定,而是体现在完整性、规范性、准确性、均衡性与时效性五个特性上。具体来说,完整性体现为要素覆盖到位、上下文可还原;规范性体现为字段类型、取值域与术语映射的一致可验;准确性依赖室内/室间质控、仪器校准与批号变更等运行证据与结果的绑定,避免非生物学误差在统计与学习中被“放大”;均衡性关注人群与疾病谱、阳性/阴性与急慢性等比例的合理分布,减少模型对机构或群体的系统性偏倚;时效性则以入库延迟与可用时窗加以度量,使数据“新鲜”的同时不牺牲质量。

为了让上述要求可实施且可评估,需要把语义与结构对齐到可计算的公共标准,例如用逻辑观察标识名称与代码(logical observation identifiers names and codes, LOINC)表达“检测了什么”,用统

一计量单位编码规范(unified code for units of measure, UCUM)统一“如何度量”,用医学系统化命名-临床术语(systematized nomenclature of medicine-clinical terms, SNOMED CT)刻画“作用于何对象”。必要时可参考 HL7 组织提出的快速医疗互操作资源规范(fast healthcare interoperability resources, FHIR),该规范是一种用于医疗健康信息交换的国际通用标准,强调模块化结构和系统间的互操作性,可通过 Observation、Specimen、Device 等资源对字段约束及其版本演进进行表达,从而在工程上直接支撑分层抽样、时序切分与评测基准构造^[5]。此外,与之配套的质量评估体系不应止于描述性检查,而是形成“定量+定性+综合”的方法学体系:编码匹配率、单位统一率、参考区间对齐率、缺失/冲突/异常比例、入库延迟、人群分布偏斜指数、结局可关联比例与时序覆盖度等指标,与元数据完备度、标注一致性、任务适配性共同使用,保证所形成的医学检验数据集既能服务多中心研究的可重复性,也能为医学检验 AI 模型的“训练-验证-测试”全链路提供边界清晰、证据充分的数据支撑。

推荐意见 2: 医学检验标准数据集是实现检验数据治理与共享、支撑 AI 模型研发与评测并推动医学检验智能化应用落地的关键基础,应围绕高质量、规范化和应用导向进行系统构建。(1B, 共识度 96.6%)

解读与证据: 医学检验结果在临床决策中具有核心地位,但当前不同医疗机构在检验数据采集、命名、存储与传输方面存在显著差异,制约了数据互认、多中心整合及 AI 模型的稳定应用^[6]。构建医学检验标准数据集,有助于统一数据口径、提升数据质量,为区域医疗协作、公共卫生应急和医学检验智能化发展提供可靠的数据基础^[7]。

在应用层面,标准数据集可服务于多类典型场景:在临床决策支持和疗效评估中,规范化检验数据可减少因数据异构导致的决策偏差^[8];在检验质量监测与方法学评价中,统一的数据结构和术语体系有助于实现客观比较与持续改进^[9];在 AI 模型训练与验证中,高质量、可扩展的数据集能够为模型提供一致的训练与评测环境,降低偏倚风险,提升模型的泛化能力与可重复性^[10]。多中心 AI 研究亦表明,严格统一的数据标准是模型性能稳定和研究成功的重要前提^[11]。

为兼顾不同应用需求,医学检验标准数据集可



在统一核心标准框架下,按应用场景构建任务相关的数据子集或扩展集^[12]。例如,面向临床决策支持、质量控制或 AI 模型训练的应用场景,可在核心数据要素基础上叠加相应的场景特有元数据,从而在保持数据兼容与复用的前提下,提升数据集的应用针对性与实际价值。

推荐意见 3: 医学检验标准数据集覆盖范围应全面,需包括血液学、生物化学、免疫学、微生物学、分子生物学等主要检验领域,并覆盖数据采集、存储、共享全生命周期。(1B, 共识度 93.3%)

解读与证据: 临床血液学检验包括血常规、凝血功能等检验项目,这些数据对于贫血、出血性疾病等的诊断和治疗具有重要意义。将其纳入标准数据集,能够为相关疾病的研究和临床诊断提供统一的数据标准。临床生物化学检验涵盖了肝功能、肾功能、血糖、血脂等多项指标,这些指标是评估人体生理功能和代谢状态的重要依据^[13]。标准化的临床化学检验数据有助于慢性疾病的监测和管理^[14]。临床免疫学检验包括抗体检测、抗原检测等,在感染性疾病、自身免疫性疾病等的诊断中发挥着关键作用。纳入该领域的的数据,可提高相关疾病诊断的准确性和一致性。临床微生物学检验涉及细菌、真菌、病毒等病原体的检测和鉴定,对于感染性疾病的诊断和治疗方案的选择至关重要。标准化的微生物检验数据有利于病原体的溯源和耐药性研究。临床分子生物学检验包括基因检测、核酸扩增等技术,在遗传病、肿瘤等疾病的诊断和预后评估中具有独特的优势,将其纳入标准数据集,能够推动分子诊断技术的规范化应用和发展。医学检验还产生多种图像数据,如微生物学涂片与染色、血涂片与骨髓细胞形态图像、病理组织切片、细胞化学染色图像等,将其纳入标准数据集,可支持图像检索与 AI 辅助分析^[15]。与此同时, AI 方法可突破单一领域的局限,通过融合血液学、生物化学、免疫学、微生物学及分子生物学等多维检验数据进行联合学习,从而建立跨检验亚专业的疾病模型,提升疾病风险预测、诊断、预后评估和精准治疗等能力,释放 AI 在多模态、跨领域临床决策中的潜能^[16-17]。

在覆盖范围上,未来应重视床旁检测(point-of-care testing, POCT)数据的纳入。POCT 具有实时性强、设备多样、环境复杂及质控难度大等特点,给标准化带来独特挑战,但将其纳入标准数据集可显著提升在急诊、基层医疗与疫情防控等场

景的适用性。未来需关注设备与接口的互通、质控信息的完整记录及标本与操作上下文的规范,宜通过推动标准化数据输出、构建 POCT 专用最小元数据集以及在 LOINC 等国际编码体系中明确映射,实现语义与兼容性的统一,从而让标准数据集更好地服务不同层级的医疗场景。POCT 最小元数据清单可包括:检测设备/方法学标识、样本类型、采样与检测时间戳、操作者/科室(或点位)标识、质控状态(如质控通过/失败、批号/校准信息)以及结果确认/上传状态等,以保障数据可追溯性与可解释性。对于急诊检验等高时效场景,应在标准数据集构建与治理中明确时间戳要素与联动规则,以支持后续模型开发与评估的真实世界适配。除常规字段外,建议至少保留采样时间、接收时间、检测完成时间、结果确认时间与结果推送/入库时间等关键节点,以便计算并监测“入库延迟”“报告延迟”等过程指标。可将“结果确认至入库延迟”作为核心及时性指标,建议参考阈值 ≤ 30 min;对危急值相关结果,建议同时记录危急值判定标识、通知对象与通知完成时间,并引入“危急值通知时效”指标: < 15 min 为“及时”, $15\sim 60$ min 为“可接受”, > 60 min 需纳入持续改进范围^[18]。同时建议明确“危急值关联规则”:将危急值标识与对应的检验项目、样本类型、临床科室/病区、通知流程与结果确认状态进行关联记录,确保数据可追溯并可用于后续质控模型与风险预警模型的训练与评估。在公共卫生应急等需要快速汇聚数据的场景中,建议在保证合规与安全的前提下,统一数据口径与最小要素集合,重点强化项目编码、检测方法学/平台、样本类型、采样与报告时间、结果判读规则(阳性/阴性/阈值)以及机构/区域标识等要素的一致性,以支持跨机构汇聚与趋势分析。

通过构建统一的医学检验标准数据集,必将为高质量医学研究、高级别临床证据产出和普惠性 AI 应用铺平道路,最终推动精准医学的发展和全民健康水平的提升^[19]。

第三部分 医学检验标准数据集构建的原则

本章提出医学检验标准数据集构建应遵循的通用原则与底线要求,明确在合规、安全与临床价值导向下实现数据标准一致、可追溯与可复用,为全生命周期构建方法提供约束框架。

推荐意见 4: 医学检验标准数据集的构建应遵



循隐私性与安全性原则。(1B, 共识度 93.3%)

解读与证据:医学检验数据属于高度敏感的健康信息,其不当使用或泄露可能对患者个人、医疗机构和全社会造成严重影响。因此,在医学检验标准数据集的构建过程中,隐私性与安全性应是贯穿始终的基本前提和底线要求^[20]。

隐私性原则强调对个人健康信息的合理控制与规范使用,其核心在于尊重数据主体的知情权与自主权,确保数据的采集、使用和共享具有明确、正当且可追溯的目的,避免超出医学检验、科研或公共卫生的合理需求范畴^[21]。医学检验标准数据集的设计,应遵循最小必要原则,在满足特定应用场景需求的前提下,合理限定数据要素范围,降低潜在隐私风险。

安全性原则强调对已采集数据的系统性保护,是保障数据完整性、可用性和可信度的重要基础。医学检验标准数据集规模大、范围广,一旦发生数据泄露或数据篡改,不仅影响 AI 系统的可靠性^[22],也可能损害临床决策安全。因此,数据集构建必须在制度流技术层面建立清晰的安全体系,确保数据在存储、应用全过程中处于受控状态。

有关医学检验数据伦理治理、安全管理及具体实施要求,将在“第五部分 医学检验数据伦理及安全”中进一步系统阐述。

推荐意见 5:医学检验标准数据集的构建应遵循临床相关性原则。(1B, 共识度 96.6%)

解读与证据:医学检验标准数据集的最终目的是为临床提供有价值的数据,不能成为孤岛数据集,一个技术上完美且标准化但临床无法使用或无法信任的数据集是无意义的。大量研究表明,缺乏临床语境的数据是导致临床决策支持系统警报疲劳、误诊和医疗错误的重要原因之一。例如,“肌钙蛋白”值升高,在急性胸痛患者中是心肌梗死或不稳定型心绞痛等心肌损伤疾病的典型标志物,但在终末期肾病患者中可能是肾脏对肌钙蛋白的代谢和清除能力减弱,导致血液中肌钙蛋白水平假性升高(即非心肌损伤引起的升高)。因此,标准数据集必须包含标本类型(如血清、血浆)、检测时间点(如入院时、症状发作后 X 小时)以及关键的临床情境信息(如主诉、诊断、用药史、肾功能状态),否则将导致严重的临床误读风险。在《急诊检验能力建设与规范中国专家共识》中明确提出某种检验项目在何种临床语境下可发布报告,表明将检验结果与特定临床场景紧密结合的实践必要性,为国内医学检

验标准数据集构建应包含哪些语境字段提供了直接依据^[23]。同时,医学检验标准数据集应能结构化地表达临床指南中的关键建议。例如,对于心力衰竭患者,数据集应能支持 B 型利钠肽和 N 末端 B 型利钠肽前体及其动态变化值的提取,并与纽约心功能分级等临床评估字段关联。美国临床和实验室标准协会指南中强调,实验室检验的价值核心体现在于改善临床结局,而构建具有结构化能力的医学检验数据集,对于评估医疗质量和开展价值医疗至关重要^[24]。

推荐意见 6:医学检验标准数据集的构建应遵循国际与国家通用标准,统一字段命名与基本定义。(1B, 共识度 96.6%)

解读与证据:医学检验数据来源广泛,涉及不同医疗机构、信息系统、检测平台及检验专业,若缺乏统一的标准,极易出现字段歧义、语义偏差等问题,严重制约数据共享及 AI 应用。因此,遵循国际与国家通用医学信息标准,建立统一的字段命名与基本定义体系,是构建医学检验标准数据集的基础性原则。

统一字段命名与基本定义的核心目的,在于实现数据语义的一致表达,使不同来源、不同时间的检验数据在合并、比较和应用时具备明确且稳定的含义,避免各机构、各系统基于本地习惯形成的数据表达方式,从源头减少数据整合过程中的歧义和误判风险^[25]。

从长远发展看,医学检验标准数据集不仅服务于当前的临床与科研需求,还需具备持续积累更新的能力。遵循成熟、开放且被广泛认可的标准体系,有助于保障数据集在不同地区、不同系统和不同应用环境中的互操作性与可扩展性,为后续的数据共享、模型建立和行业应用奠定坚实的保障基础。

有关具体标准的选择、字段映射策略及工程实现路径,将在“第四部分 全生命周期视角下的医学检验数据集构建方法”中进行系统阐述。

推荐意见 7:明确并记录数据采集的医疗机构、设备型号、检验方法等元信息,统一度量单位,优先采用国家法定计量单位,以保证多中心数据可整合与可溯源。(1B, 共识度 93.3%)

解读与证据:多中心数据是 AI 建模与循证研究的基础,但方法学差异、设备差异与度量单位不统一将导致系统性偏差,影响模型泛化与研究结论的可比性^[26]。强烈建议在实验室信息系统



(laboratory information system, LIS)层面即实现度量单位的统一和 UCUM 编码的采用,并强制记录设备型号、试剂批号、校准溯源性等关键元信息,在系统层面实现单位自动校核与转换,避免如血糖单位 mmol/L 与 mg/dl 混用导致的偏差与错误,进而影响数据的可用性。国家标准 GB 3100—1993《国际单位制及其应用》明确要求,医学检验结果使用全球法定单位国际单位制(Système International d'Unités, SI)体系,国际计量局(Bureau International des Poids et Mesures, BIPM)的 SI 手册明确单位定义、派生单位与量纲关系,为检验结果的一致表达与跨机构可比提供依据。元数据中记录的检验方法、仪器型号、实验室编号等信息,应参照《医疗器械唯一标识系统规则》(国家药品监督管理局 2019 年第 66 号公告)中关于设备标识与追溯的要求。欧盟的体外诊断医疗器械条例(In Vitro Diagnostic Medical Device Regulation, IVDR)同样强调实验室数据的可追溯性和单位一致性。世界卫生组织实验室质量管理体系(World Health Organization Laboratory Quality Management System, WHO LQMS)将“过程方法”和“以患者为中心”作为实验室质量管理与可追溯性的核心,要求在检验前、中、后过程中明确样本、方法、设备、环境、人员等与信息的质量控制;数据集中需完整记录来源机构、仪器与试剂信息、方法学与性能声明、标本类型与前处理、校准与质控记录等,以支持结果的溯源与解释^[27]。在互操作实现上,HL7 FHIR 的 Specimen 与 Device 资源为标本与设备元数据提供标准结构,Observation 资源可绑定方法学、单位、参考区间与异常标识;在仪器层面,国际医疗器械监管机构论坛的医疗器械唯一标识(unique device identification, UDI)指南为唯一器械标识提供国际框架,便于在数据集内记录与追踪具体设备与版本差异^[28]。规范多中心元数据并非单纯“记录信息”,更是保证结果可比性的必要措施:纳入方法学性能(线性范围、检出限、精密度)、校准物与质控批次、环境与维护记录,才能解释同一分析物在不同中心的差异,为后续标准化换算或分层分析提供依据。通过引入“实验室 ID”“仪器品牌”“检测方法”等元字段,可在模型训练阶段进行分层或归一化校正,提高泛化能力。美国食品药品监督管理局(Food and Drug Administration, FDA)在其基于 AI/机器学习的软件即医疗器械指导原则中也指出,多源异构数据统一标准是算法稳健性的核心^[29]。建议在数

据集规范中明确字段必填项与校验规则(如单位必须为 UCUM 编码、仪器需有 UDI/型号与软件版本、方法学需选自受控术语表),并提供映射与转换指南及示例,保障不同中心数据的高质量整合。

推荐意见 8: 建立严格的数据采集、审核与清洗流程,明确缺失值、异常值处理规则,提升数据的准确性、完整性和可复用性。(1B, 共识度 93.3%)

解读与证据: 数据质量直接决定 AI 应用与科研的有效性。WHO《全球数字健康战略(2020—2025)》指出,医疗数据质量控制应贯穿于采集、传输、分析全流程^[30]。缺失值、异常值、录入错误及时间戳不一致等可导致系统偏差,降低预测模型性能。研究显示,随着数据集缺失率的增加,模型的判别能力即受试者工作特征曲线下面积(area under the curve, AUC)可能显著下降^[31]。数据质量控制应包括:(1)采集端校验,防止录入错误;(2)审核端比对,通过逻辑规则核查异常结果;(3)清洗端建立统一缺失值处理策略(如分组插补、标记缺失等)。对于缺失值的处理规则,尤其是可能具有临床意义的缺失(如因标本量不足无法检测),必须由检验医师和临床医师共同参与制订,避免单纯依靠统计方法抹除重要临床信息。此外,需保留“数据溯源链”,确保任何结果可回溯至原始记录。欧盟《通用数据保护条例》(General Data Protection Regulation, GDPR)第 5 条明确要求数据应“准确且及时更新”;中国《电子病历系统应用水平分级评价管理办法(试行)》也规定数据准确、完整、可追溯。引入质量审计、自动规则校验和人工复核结合机制,可显著提升数据集可靠度。在证据综合与决策支持层面,英国国家卫生与临床优化研究所(National Institute for Health and Care Excellence, NICE)的诊断技术评估方法将诊断准确性、临床有效性、经济学证据整合为框架,提示数据集需要满足统计与决策分析的可计算性,例如纳入参考区间来源、方法学性能、关键协变量与数据质量标记,支持后续的敏感性分析与分层评估^[32]。此外,测量不确定度的量化对于结果解释与跨中心比较至关重要,计量学指南联合委员会的《测量不确定度表示指南》(Guide to the Expression of Uncertainty in Measurement, GUM)为识别、量化与合成不确定度提供了通用方法学,建议在数据集中保留不确定度或测量误差范围(如标准不确定度或置信区间)的记录,便于临床使用者与建模者选择利用^[33]。具体实施建议包括:制定统一的数据字典与校验规则



(单位、取值范围、时间格式、标本一致性)、建立双人审核与自动规则校验机制、记录质量事件与修订历史、对缺失值采用场景化策略(区分随机缺失与非随机缺失)、对异常值进行方法学与临床双重判定(如结合设备质控、危急值与重复检验信息)。通过“过程质量+透明元数据+不确定度量化”的三位一体模式,可显著提升数据的可靠性、可比性与复用价值。

第四部分 全生命周期视角下的医学检验 数据集构建方法

本章核心目标在于系统阐述医学检验标准数据集在构建、质量控制及持续更新过程中的全生命周期实施方法,将构建原则转化为可操作的技术与管理路径。医学检验标准数据集的全生命周期构建框架见图 1,其中伦理与安全要求贯穿全过程。需要说明的是,第三部分所提出的构建原则旨在概括医学检验标准数据集构建应遵循的通用约束与价值取向,而本部分从数据采集、清洗与标注、数据质量控制与评估以及数据更新与迭代等不同阶段,具体展开数据集构建的实施方法。在这一框架下,构建原则与具体方法之间并非一一对应的线性关系,而是呈现出多对一或一对多的交叉对应特征:同一原则往往贯穿多个构建环节,而单一方法步骤亦需同时满足多项原则要求。基于此,本部分各节内容均在第四部分构建原则的总体框架下展开,旨在从全生命周期角度系统落实相关原则要求。

一、检验数据的采集、清洗和标注

推荐意见 9: 医疗机构在检验设备接口改造和新增设备采购时,宜优先采用符合主流 HL7 标准的数据输出接口,并建议采用 LOINC/SNOMED CT 编码体系统一检验项目、方法与单位等字段命名,实现多源设备数据的结构化语义整合。(1B, 共识度

96.6%)

解读与证据:检验数据采集是医学数据生命周期的基础环节,其标准化程度直接决定后续数据整合、分析与 AI 建模的可行性与有效性。当前医疗机构普遍面临检验设备品牌多样、接口不一、数据格式异构的挑战。这种异构性不仅增加数据提取与清洗的技术复杂度,也导致数据在系统间互通困难,形成“数据孤岛”,制约多中心研究、真实世界研究及 AI 预测模型的开发与应用^[34]。

从计算机科学角度实现设备数据标准化输出是构建可扩展、互操作医疗信息系统的重要前提。在接口标准选择上,目前国际上已形成以 HL7 v2.x ORU 消息与 HL7 FHIR R4/R5 Observation 资源为主、以 ASTM E1381/E1394 作为部分分析仪底层通讯协议并通过中间件映射到 HL7/FHIR 的技术格局,其应用有助于降低系统集成成本,提升数据流动效率^[35-36]。同时,采用 LOINC 与 SNOMED CT 作为语义标准,可实现检验项目名称、方法、单位等关键信息的无歧义表达,为后续的自然语言处理与机器学习任务提供更高质量输入^[37-38]。

从统计与建模角度看,数据标准化是保障样本可比性、减少偏倚的重要环节。若不同机构或设备使用各自命名体系,即便同一检验项目也可能因命名不一致而被误判为不同变量,造成合并分析中的信息损失或错误归类。既往多项真实世界研究提示,LOINC 编码的系统性应用可显著提升变量对齐的准确率,进而增强模型的泛化能力^[39-40]。

此外,国际大型队列项目如美国“All of Us”研究计划^[41]、英国生物银行(UK Biobank)^[42]普遍采用 LOINC 与 SNOMED CT 进行检验数据编码,为跨学科、跨机构合作提供了可复用的基础。因此,本推荐意见在结合国内实践的基础上参考国际共识,旨在推动我国医学检验数据从采集源头逐步具备“FAIR”(可发现、可访问、可互操作、可重用)原则要求的基本属性。

推荐意见 10: 在多中心或跨机构构建医学检验数据集时,宜在跨机构交换层优先采用基于主流 HL7 标准的接口规范,结合 LOINC 编码及 ISO 15189 数据模型实现结构化语义约束;对于底层仅支持传统协议的老旧设备,建议厂商提供符合网络安全与国密算法要求的协议转换中间件,以支持“采集即标准

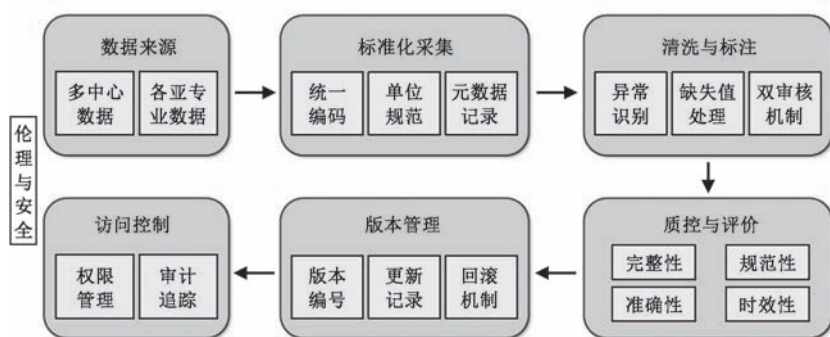


图1 医学检验标准数据集全生命周期构建与治理框架



化”,提升数据在源头的互认性、可溯源性 AI 训练可用性。(2C,共识度 90.0%)

解读与证据:跨机构、跨省域环境中,LIS 品牌和设备接口长期呈现异构特征,字段命名、计量单位、参考区间等要素在不同机构间错配较为常见,进而影响检验结果互认与 AI 模型跨中心部署。国家卫生健康委员会《医疗机构检查检验结果互认管理办法》(2022)提出“统一数据标准”,在此背景下,为增强跨机构交换的一致性与可治理性,数据采集与交换层面可优先采用 HL7 FHIR R4 Observation 资源作为消息骨架,优先使用 LOINC 编码标识检验项目并按 UCUM 规范计量单位,同时参考 ISO 15189:2022 数据模型对参考区间、质控与设备信息进行约束;对无法直接输出的老旧设备,可通过厂商提供的可配置 ASTM 1394 或 HL7-to-FHIR 转换中间件实现协议映射与语义对齐,并在条件允许时通过网络安全等级保护 2.0 三级测评、支持国密算法(SM2/SM3/SM4),以降低数据交换与存储的安全风险。

现有研究与实践为上述路径提供一定证据支持:FHIR 作为一种以资源为核心的互操作性标准,在健康研究与多源数据整合中被广泛采用。研究认为,将 FHIR 与标准化术语体系如 LOINC 结合,有助于提升实验室及临床数据的语义一致性和跨系统可交换性,并在多中心研究场景下减少数据映射与清洗的复杂度。相关证据主要来源于多中心实践研究、原型系统和真实世界应用案例,整体结果支持基于 FHIR 的统一数据接口在改善数据异构性和提高数据复用效率方面具有积极作用^[39];同时,ISO 15189 作为医学实验室质量与能力的国际通行标准,明确强调检验结果的可追溯性、可比性及过程一致性,对检验数据的结构化记录和质量元数据管理提出了制度性要求^[43]。从技术层面出发,FHIR Observation 以 RESTful JSON 为载体,契合我国《“十四五”全民健康信息化规划》提出的“互联网+医疗健康”微服务架构。综上,基于 HL7 FHIR、LOINC 与 ISO 15189 的“采集即标准化”策略可能是当前破解检验数据孤岛、推进 AI 临床转化的重要路径。

推荐意见 11:检验数据清洗规则宜由检验人员与临床医师共同制订,建立“技术-临床”双审核机制。数值型数据标注建议尽可能关联最终诊断[国际疾病分类(International Classification Of Diseases, ICD)-11 编码]及临床决策节点(如治疗启

动/调整时间等)。(2C,共识度 93.3%)

解读与证据:数据清洗与标注是从原始数据转化为可分析、可建模高质量数据集的核心环节。传统做法多由数据科学家或统计人员主导,主要依赖箱线图、Z-score 等统计规则识别异常值与缺失值。然而,医学数据具有高度情境依赖性,单纯依赖统计规则极易误删具有临床意义的“真异常”,如肿瘤标志物的急剧升高、危急值波动等,从而影响对关键临床信号的识别与建模效果^[44-45]。对于后续识别出的异常值误判情况,应建立复核与修正流程,保留原始数据记录及处理痕迹,确保数据清洗过程可追溯、可审计。

从医学检验专业视角看,检验人员对检测方法学原理、干扰因素及临床意义的把握,有助于在清洗阶段对溶血、脂血或药物干扰等造成的异常数据进行识别与标记,而非简单剔除^[46]。同时,临床医师对患者整体病情、诊疗方案与时间线更为熟悉,可辅助判断检验结果是否与诊断相符、是否提示疾病转折点。因而,“检验-临床”双审核机制有助于提升数据的临床解释性与生物学合理性。在实施层面,可由检验人员制订异常识别规则并进行系统初筛,对疑似异常数据由临床医师结合诊疗情境进行复核;对确认的误判情况,通过规则修订与记录留痕形成持续改进机制。

从统计学建模角度来看,标注质量直接影响监督学习模型的性能。若仅标注检验数值而缺乏对应的临床结局(如 ICD-11 诊断)与关键时间节点(如治疗响应、复发时间),将导致模型无法学习到有预测意义的动态模式^[47-48]。在脓毒症预测、急性肾损伤等时间敏感性疾病中,标注“临床决策节点”(如抗生素使用时间、肾脏替代治疗开始时间),可极大增强模型对前驱信号的识别能力^[49-50]。

从 AI 实践来看,数据标注的规范性与一致性是模型可复现性的基础。ICD-11 作为国际统一的诊断术语体系,有助于提高标注一致性^[51]。结合临床路径对检验序列进行阶段标注(如术前、化疗后、随访期),也有助于构建更具泛化能力的阶段感知预测模型^[52]。国际上多个标杆性医学 AI 项目(如 MIMIC-IV、eICU)已采纳类似跨学科清洗与标注策略,并公开发布其数据治理流程,极大促进了第三方验证与模型迭代^[53-54]。此外,在医学检验标准数据集构建过程中,针对人工标注成本高、专业依赖性强等问题,可在符合伦理与安全要求的前提下,探索性地引入生成式 AI 等新技术作为辅助标注工



具。例如,生成式 AI 可用于辅助检验项目或结果的初步标注,提示潜在标注一致性问题,从而提高标注效率并降低人工负担。需要强调的是,此类技术仅作为辅助工具使用,最终标注结果仍应由专业人员审核确认,以确保标注的准确性与可靠性。因此,本推荐意见旨在通过协同治理提升数据质量与应用价值,并为医学 AI 生态的规范发展提供基础条件。

推荐意见 12: 对所有进入数据集的检验缺失值,应按缺失机制进行分类并赋码,同时在数据表内保留“缺失原因”字段,并随数据版本同步更新。(2C, 共识度 90.0%)

解读与证据: 跨机构数据缺失率常高达 30%~50%, 原因包括仪器漏检、标本量不足、网络中断或临床取消等。若沿用传统均值插补或末次观测值结转法等处理方式,可能掩盖真实临床情境,影响危急值预警准确性,并进一步影响 AI 模型的可信度与可解释性。国家卫生健康委员会《医疗机构检查检验结果互认管理办法》(2022)强调“完整记录缺失原因并保证可追溯”。在国际上,美国国家研究委员会指出,明确缺失机制有助于避免“错误修复”带来的因果扭曲。在国内研究中,新疆维吾尔自治区 2024 年开展的真实世界队列研究(包含 10 164 名受试者,涉及 37 个变量)显示,采用均值插补等方法可能会显著削弱模型区分能力即 AUC 降低,相较之下,保留缺失信息可在一定程度上反映缺失机制,从而维持模型 AUC 的稳定性^[55];也有研究从国内不同地区的 4 家三级医院收集数据,采用 7 种机器学习算法构建预测模型,其中人工神经网络模型表现出优异的预测性能,并且基于沙普利加法解释(SHapley Additive exPlanations, SHAP)对该模型的可解释性进行分析^[56];SHAP 解释提示的“UNTEST”(设备未测)等标签本身可作为负向风险特征,有助于快速识别漏检风险。在实施层面,建议在 LIS 或中间件中增设字段“missing_reason”,通过下拉菜单规范分类,并同步写入 HL7 FHIR Observation.extension,以减少跨省共享时语义丢失;同时结合国家信息安全等级保护 2.0 三级要求,采用国密算法(SM2/SM3/SM4)加密存储,以符合《数据安全法》要求。值得一提的是,在遵循既有缺失值分类与标识原则的基础上,可探索性地借助生成式 AI 等技术,辅助开展缺失模式的分析与分类,以支持对缺失机制的识别与解释。在特定研究或方法学评估场景下,生成式技术亦可用于构建模拟

数据,以验证缺失机制对数据结构和分析结果的潜在影响。综上,缺失原因分类赋码既符合我国检验数据治理现状,也获得国内外高质量研究与政策支持,是兼顾准确性、可解释性和经济性的可行路径。

二、数据质量控制与评估

推荐意见 13: 医学检验标准数据集应在全生命周期(采集-预处理-存储-共享-再利用)建立系统化的数据质量管理与评价体系,明确质量目标、评价维度、责任分工与反馈机制,以持续支持科研、临床及 AI 模型开发的质量要求。(2C, 共识度 96.6%)

解读与证据: 医学检验标准数据集作为医学检验行业的通识数据集^[57],常用于训练或评估具备医学检验行业能力的 AI 模型(如病理与细胞形态学、基因与分子诊断、检验质控与流程优化、报告解读、智慧实验室等)。医学检验标准数据集是医学检验行业 AI 模型的“核心基石”,其质量水平直接影响 AI 模型性能和实际应用效果。开展系统、规范的数据集质量评价,一方面有助于判断数据集是否达到“高质量”要求,另一方面也可通过“以评促建、以评促用”推动数据集持续优化,从而全面提升检验数据资源的应用价值。因此,制定统一的数据集质量评测标准,并构建覆盖全生命周期的质量评价工作体系,对于提升优质医学检验数据供给与流通、强化医学检验 AI 模型训练的数据支撑具有重要意义^[58]。

为提升数据集质量评价工作的系统性、规范性以及可溯源性,可将实施路径概括为三个阶段,各阶段既可独立实施,也可嵌入现有质量管理体系中分步推进,以增强不同实验室环境下的可操作性。

1. 评估准备阶段:在正式开展质量评价前,明确医学检验标准数据集的基本信息(包含数据模态、数据层次、应用场景、数据采集要求、数据预处理要求、数据标注要求、元数据、数据更新、数据多样性等要素)^[59]、应用场景(AI 模型训练与验证、质控检测、行业标准制定等)及质量评价的目标,并配备相应评估团队、评价工具和数据支撑环境。

2. 体系构建与实施阶段:质量评价指标体系包括说明文档、数据质量、模型应用三个评估维度^[60]。其中,说明文档完整性强调数据集需具备包含基本信息、内容特征、建设过程及应用说明的完整文档;数据质量合规性关注数据结果完整性、安全性、格式和标注规范、内容真实性、清洗过程可追溯性、内容一致性等;模型应用场景适用性则强调数据集在目标场景下对 AI 模型性能的支持程度(如数据内



容多样、结构完整、标准准确)。以上三个维度的质量指标可通过计算方法进行量化,并兼顾数据的描述规范性、本体质量和应用效果。

3.综合评价阶段:基于质量评价指标结果,结合多维度指标体系进行加权汇总,形成量化的质量评分与等级划分,并形成综合评价报告,用于指导数据集维护迭代与应用推广,确保医学检验数据集在全生命周期过程中发挥可靠的支撑作用。

推荐意见 14:在全生命周期质量管理与评价框架下,医学检验标准数据集应建立基于完整性、准确性、一致性、及时性、可靠性、安全性等核心维度的量化指标体系,并根据不同医学检验场景和应用需求合理设置各维度权重,形成综合的数据质量评分体系,可鼓励采用数据质量评分卡作为多中心统一的质量度量工具。(2C,共识度 93.3%)

解读与证据:全面的数据质量评估维度能够更综合、客观地反映医学检验数据集的质量状况。本共识建议从完整性、准确性、一致性、及时性、可靠性、安全性六个核心维度出发对数据集进行量化评估,并给出推荐性指标与目标阈值(表 2)。不同质量维度反映数据在不同层面的可信度与安全性,可通过标准化定义、指标体系与技术方法实现科学评估与动态监控。例如,完整性可通过字段覆盖率(field coverage rate, FCR)、缺失率(missing rate, MR)及基于数据库查询或编程分析工具(如 SQL 脚本、Python/Pandas)的缺失可视化进行评估;一致性可通过单位与编码一致性矩阵、相关一致性系数(concordance correlation coefficient, CCC)及累积和控制图(cumulative sum control chart, CUSUM)实现

漂移检测进行跨机构和跨时间分析;安全性可依据 GB/T 39725—2020 等国家标准,对权限管理、日志留存和入侵检测等进行系统审计。表 2 所列阈值为参考国内外文献与专家经验提出的推荐目标值,各中心可结合自身信息化成熟度与数据质量现状分阶段达标,并在质量评价报告中记录实际达成情况及改进计划。

不同应用场景对各维度的重要性权重通常有所侧重,例如,用于临床决策支持的实时预警模型对及时性和可靠性的要求更高,用于长期流行病学分析的数据集则更强调完整性与一致性。因此,通过德尔菲法开展专家咨询,根据专家打分与意见对指标进行增减、合并或修改,并采用重要性评分法确定各维度及其下属指标的权重,形成可解释的综合评分体系,即数据质量评分卡。在技术实现层面,可参考《OHDSI Data Quality Dashboard》,开发 R/Python 开源脚本,实现对 HL7-FHIR、LOINC、SNOMED CT 等标准的自动映射与规则运行,便于多中心自动化部署与结果可视化展示。

除上述质量指标外,还应考虑面向应用的“临床适配性”过程指标,用于评估数据集对真实业务流程的支撑程度。临床适配性指标可包括:(1)临床场景覆盖率(目标场景所需关键字段覆盖程度);(2)关键流程时效指标(如结果确认至入库延迟、报告延迟);(3)危急值数据完备性与通知时效(记录是否齐全、通知时效分档);(4)可行动信息比例(与临床决策相关字段/元数据完备度)。建议将上述指标与技术质量指标并行呈现,形成“技术质量+过程适配性”的双维度评估框架,为后续模型训练、评

表 2 医学检验数据质量控制与评估体系

质量维度	定义与意义	评估指标与方法	合格阈值/要求	备注
完整性 ^[61]	检验申请-结果-随访全过程关键字段齐全程度	FCR、MR;基于数据库查询或编程分析工具进行批量统计与缺失可视化	FCR≥98%、MR≤2%	以 LOINC/SNOMED CT 编码为必填字段依据
准确性 ^[62]	数据真实反映检验实体的程度	金标准比对法、多中心交叉验证法(偏倚分析)	AS≥98%、CV≤5%	建议≥5%随机样本抽检
一致性 ^[63]	数据格式、单位、编码的统一性	格式、单位、编码一致性矩阵分析;CCC 计算、CUSUM 漂移检测	CCC≥0.95	支持跨中心、跨时间一致性评估
及时性	数据采集、报告、更新的时效性	时间戳追踪法;计算采集延迟、报告周期、更新频率	延迟率≤5%	每月生成及时性分析报告
可靠性 ^[64]	数据重复性与临床逻辑一致性	室内质控重复测量、临床医师抽样复核	CV≤5%;抽样一致率≥95%	每月抽取≥100份异常结果复核
安全性 ^[20]	数据保密性、完整性与可用性	安全审计(权限、日志)、技术检测(IDS、DLP)	满足 GB/T 39725—2020 等要求	日志留存≥6 年、泄露后 2 h 启动应急响应

注:FCR 为字段覆盖率;MR 为缺失率;CCC 为相关一致性系数;CUSUM 为累积和控制图;IDS 为入侵检测系统;DLP 为数据防泄漏;AS 为准确性得分;CV 为变异系数;LOINC 为逻辑观察标识名称与代码;SNOMED CT 为医学系统化命名-临床术语



测与上线前验证提供基础依据。

推荐意见 15: 为保障既定的数据质量要求在全生命周期内得以落实与持续改进,建议构建多层次、分级负责的数据质量三级审核机制,通过实时监测、定期抽检及集中回溯审核形成由点到面、由局部到全局的闭环质量管理体系,以提升数据集在全生命周期内的稳定性与可靠性。(2C, 共识度 93.3%)

解读与证据: 在日常运行中,仅依靠离线统计或单次质控可能难以及时发现数据质量问题^[65],容易造成隐性错误累积,影响后续分析与模型训练。为此,建议建立覆盖“实时-周度-月/季度”三个频次的三级审核机制。

1. 一级(实时审核): 由 LIS 或数据中间件内置规则引擎完成,规则库数量宜不少于 150 条并定期更新(如每季度更新一次);重点监控异常值、离群值、单位换算错误、非法编码、关键字段缺失等问题,对严重异常可自动触发拦截、复核或告警流程。

2. 二级(周度抽检): 由各中心数据质控小组按不低于 2% 比例随机抽检,重点核查缺失值填补逻辑、离群值处理记录等;抽检结果可通过可视化仪表盘反馈给实验室负责人和信息部门,用于指导规则库调整和流程优化。可自动触发拦截、复核或告警流程。

3. 三级(月/季度集中审核): 由多中心联合质控管理委员会组织,采用集中式数据湖与分布式计算框架进行全量回溯评估,重点关注各中心在完整性、一致性、及时性、安全性等方面的长期表现,识别系统性偏差或结构性问题。审核报告宜在 10 个工作日内完成并向所有共建单位发布,必要时形成整改计划与时间表。

三、检验数据的更新与迭代

推荐意见 16: 医学检验标准数据集及其配套术语字典应建立版本管理与变更记录机制,以支持数据集动态反映最新的临床实践和实验室技术发展,并通过明确版本编号、更新频率和变更流程等方式,保障变更的可追溯性和可回滚性。(1C, 共识度 96.6%)

解读与证据: 医学检验数据是医疗信息化和 AI 发展的重要基石,其应用范围覆盖临床诊疗、科研探索、质量控制和公共卫生决策。随着液体活检、单细胞测序以及质谱检测等技术不断涌现,传统检测项目的参考区间、方法学和临床应用范围也在循证医学证据的积累中持续修订。同时,疾病

谱变化、诊疗指南更新与政策导向的调整亦会推动检验数据在临床中的价值和使用方式发生变化。在此背景下,若缺乏系统的版本管理和更新机制,标准数据集可能滞后于实际应用,造成科研与临床间的“错位”,甚至可能影响 AI 模型的准确性与可推广性^[66]。

国际上多项成熟的医疗信息标准均建立了完善的版本管理体系。以 HL7 FHIR 为例,其核心资源每次更新都会发布明确的版本号,并在发布文档中标注每一次迭代的新增、修改和废弃内容,从而保证用户能够清晰地追踪数据标准的演进过程^[67]。美国卫生信息技术国家协调办公室在《美国互操作性核心数据集》(United States Core Data for Interoperability, USCDI)中提出核心数据集需定期更新,以纳入新的数据元素并淘汰过时内容,更新周期通常为 1~2 年^[68]。LOINC 术语体系(当前版本 2.81)每年 2 月与 8 月例行更新两次^[69],欧洲的体外诊断相关标准亦通过定期版本发布,促进数据标准与诊断产品注册及临床应用保持一致^[70]。这些经验提示,版本管理不仅是技术问题,更是保障数据标准权威性和生命力的关键环节。

国内方面,国家卫生健康委员会发布的《电子病历共享文档规范(WST 500)》和《国家健康医疗大数据标准、安全和服务管理办法(试行)》均明确提出“数据元需随政策、技术和疾病谱发展进行动态更新”^[71-72]。《ISO 15189:2022 医学实验室质量与能力的要求》也强调实验室在质量管理体系中应建立文档控制和更新机制,保证所使用的标准与当前最佳实践保持一致。《卫生健康信息数据元标准化规则》和《卫生健康信息数据集元数据标准》等行业标准也对数据元的版本属性、变更记录等作出了具体规定,为医学检验标准数据集的版本管理提供了直接的制度依据。

医学检验标准数据集的版本管理具体实施层面建议重点关注以下 4 个环节。

1. 数据集整体版本编号与更新频率: 建议采用国际通用的主版本号+次版本号(如 v1.0、v1.1)的方式标识更新范围与深度。常规更新可每 1~2 年开展一次,用于纳入新检验项目、调整参考区间或优化数据结构;如遇重大公共卫生事件或新兴检测项目快速应用时,可启动加速更新机制。

2. 术语库与编码映射的版本管理: 建议同步记录 LOINC、SNOMED CT 及本地检验术语库的版本号,标注新增、修改和废弃条目,并维护字段值域和



编码映射的变更历史(如新增 LOINC 映射、参考区间调整、方法学变更等)。变更记录宜结构化保存,经伦理与质量审查后发布更新说明,并与 LIS/接口系统同步验证,必要时开展回归测试。

3. 更新日志与历史版本存档:每次数据集更新建议同步发布更新日志,明确新增、修改和废弃内容及生效时间,方便临床和科研人员理解并及时调整数据应用策略。所有历史版本宜长期保存,并建立可对比和可追溯机制,以支持长期纵向研究和模型复现。

4. 变更通知与用户培训:更新内容应通过权威渠道及时发布,并配套用户培训和宣贯,便于一线使用人员能够快速理解并正确使用新版数据集与术语映射。

总体而言,在遵循国家标准总体框架的前提下,借鉴 HL7 FHIR、USCDI、LOINC 等国际经验,建立例行更新与应急更新相结合的版本管理机制,使数据集在保持稳定性的同时具备足够的敏捷性,更好地支撑检验结果互认、真实世界研究和医学 AI 模型的迭代优化。在必要情况下,应具备回退至既往稳定版本的能力,以降低版本更新可能带来的系统性风险。

推荐意见 17:建议在检验数据集更新迭代过程中引入临床需求和真实世界数据反馈机制,以增强数据集对医学前沿问题和临床实际应用需求的响应性与适配性。(2C, 共识度 93.3%)

解读与证据:医学检验标准数据集不仅服务科研资源,也应面向临床应用和公共卫生管理。如果更新主要停留在技术和标准层面而临床需求纳入不足,数据集的应用价值可能受到限制^[73]。例如,近年来免疫检查点抑制剂应用广泛,随之衍生的伴随诊断检测项目不断增加,若数据集更新不能及时覆盖,就会制约其在临床实践中的规范应用^[74]。国际上, FDA 在紧急使用授权机制下,通常会较快吸收市场和临床反馈,如新型冠状病毒检测试剂紧急使用授权页面所示,可及时纳入新兴检测项目和临床应用需求^[75]。在我国,国家药品监督管理局近年来加强对新检验项目注册与真实世界数据的收集,并推动将这些数据纳入国家监管数据库^[76]。因此,医学检验标准数据集的更新不宜仅局限于标准制定者的视角,可通过问卷调查、临床需求调研、真实世界数据挖掘与反馈闭环等方式,将临床一线需求和科研前沿成果同步纳入更新框架,从而持续提升数据集的适用性与应用价值。

推荐意见 18:建议根据医学检验项目的重要性、临床应用范围及新证据的涌现速度,建立分级管理的数据更新频率与优先级设置机制,以兼顾标准数据集更新的科学性、可操作性和资源合理性。(1C, 共识度 93.3%)

解读与证据:医学检验项目的更新需求差异显著。部分常规项目如血常规、电解质检测,其临床应用和方法学已较为成熟,更新频率相对较低;而分子诊断、肿瘤标志物及新型免疫学检测常随研究进展和指南修订快速演进,更新需求更为迫切^[77]。美国 USCDI 采用“核心集+扩展集”的分级模式:核心数据集更新周期相对固定(每 1~2 年),扩展数据集则允许更快地纳入新项目^[68]。欧洲 IVDR 也根据风险等级分层,对高风险 IVD 项目提出更高的更新频率^[69]。结合我国实际,建议建立“三层分级+动态调整”机制:(1)高优先级项目(如国家重点疾病防控相关项目、新兴高风险检测)建议每年评估;(2)中优先级项目(如常规生化、免疫学检测)建议每 2~3 年更新;(3)低优先级项目(如科研探索性项目),可根据需求适时更新^[78]。通过分级更新与优先级管理,有助于在有限资源条件下实现“重点优先、循证驱动、动态迭代”,提升标准数据集更新的可执行性与可持续性。

第五部分 医学检验数据伦理及安全

本章系统阐述了医学检验标准数据集在伦理合规与安全治理方面的关键要求,明确从数据收集、处理、使用到共享的风险控制路径,保障数据利用在合法合规、可审计、可追溯前提下开展。

一、医学检验数据伦理相关

推荐意见 19:医学检验标准数据集相关研究的数据收集阶段,应严格遵循《涉及人的生命科学和医学研究伦理审查办法》等国家相关法律法规开展伦理审查,其中非人类遗传资源相关研究(含单中心、多中心、国际合作等)需按对应法规要求完成层级化伦理审查;人类遗传资源相关研究需在通过机构伦理委员会审查的基础上,额外完成人类遗传资源管理相关专项伦理审查,确保符合《人类遗传资源管理条例》及实施细则的强制性要求。(1A, 共识度 93.3%)

解读与证据:所有涉及构建和使用医学检验标准数据集的研究或项目,在启动前及实施过程中的关键节点,必须接受项目所在机构的伦理委员会或



一个独立的、具备医学检验数据伦理审查专长的区域或国家级伦理委员会的审查与批准。必须确保伦理评估的全面性、独立性和专业性,尤其关注检验数据的特殊性(如高度敏感性、潜在遗传信息、诊断价值等)。

依据《涉及人的生命科学和医学研究伦理审查办法》要求,二级以上医疗机构及设区的市级以上卫生机构必须设立伦理审查委员会。对于多中心合作或商业合作项目,如机构伦理委员会审查存在专业能力不足或利益冲突问题,则需委托区域伦理审查委员会进行独立审查。区域伦理委员会由省级卫生健康主管部门管理,具备更广泛的专家资源和跨机构视角,能有效统一审查标准,提高审查效率。

《赫尔辛基宣言》强调涉及人类受试者的研究需接受独立伦理审查^[79]。《涉及人的生物医学研究伦理审查办法》要求生物医学研究需经伦理委员会审查批准。《个人信息保护法》对处理敏感个人信息(健康信息属于此类)提出了更严格的合规要求,独立监督是重要保障。国际实践(如欧盟 GDPR 下针对健康数据处理的严格要求)普遍要求独立的数据保护影响评估和伦理审查。

医学检验数据的误用或泄露风险极高,层级化伦理审查能最大程度降低伦理风险,增强公众信任。层级化伦理审查机制的适用场景及要求见表 3。

在检验数据领域,层级化伦理审查机制需特别关注以下 3 个方面。

1. 数据来源合法性:验证生物样本或健康记录的获取是否符合《人类遗传资源管理条例》及知情同意规范。

2. 跨中心数据同质化:确保不同机构检验方法、仪器平台差异不影响数据可比性,需提供标准化操作程序及质控证明。

3. 商业利益冲突管理:企业资助项目需公开经费来源,数据使用权与成果转化收益分配方案需经双重审查批准。

推荐意见 20: 检验数据集构建必须遵循隐私

优先设计原则,采用匿名化/去标识化技术处理敏感信息。数据集发布前需通过重新识别风险评估,确保无法复原至个人身份信息。(1B, 共识度 93.3%)

解读与证据:《个人信息保护法》将医疗健康信息定义为敏感个人信息,要求处理此类数据前需取得单独同意并采取严格保护措施。检验数据集构建需区分两类技术:(1)匿名化:通过删除直接标识符(姓名、身份证号)及聚合分类变量(如将精确年龄转换为年龄组),使数据无法关联到特定个体且不可复原。符合《个人信息保护法》匿名化标准的数据可免除伦理审查。(2)去标识化:保留数据可追溯性(如使用加密代号),需结合访问控制、审计日志等技术手段,此类数据仍属个人信息范畴,需持续伦理监督。

检验数据的特殊挑战在于:(1)高维度可识别性:某些检验组合(如罕见病特异性生物标志物)本身具有身份识别性,需通过数据抑制或噪声注入降低重识别风险。(2)纵向数据关联:长期随访的检验结果可能通过时间模式识别个体,需制定动态匿名化策略。

因此实施要点在于建立数据分级分类制度(表 4),按风险等级匹配保护措施。值得注意的是,去标识化是通过技术或管理手段移除或替换可直接识别个体身份的信息,在特定条件下仍可通过授权实现追溯,而匿名化则是指对数据进行不可逆处理,使其无法再识别或关联至特定个体。

推荐意见 21: 涉及儿童、孕产妇、精神障碍患者等特殊人群的检验数据,需执行额外保护措施。(1B, 共识度 93.3%)

解读与证据:国际伦理准则与《涉及人的生命科学和医学研究伦理审查办法》均要求对弱势群体给予特殊保护。在检验数据收集中,针对不同特殊群体的风险特点,需采取相应的保护性措施:(1)儿童群体:遗传代谢病筛查数据可能影响未来保险权益,需设置数据访问权限与留存时限。例如新生儿遗传病筛查阳性结果用于科研时,需删除家庭住址等间接标识符。(2)孕产妇群体:产前检验数据(如

表 3 层级化伦理审查机制的适用场景及要求

项目类型	机构伦理审查委员会审查要求	区域伦理委员会/专项伦理审查要求	审查重点
单中心非敏感数据研究	必须审查	选择性审查	数据匿名化措施、知情同意流程
多中心研究	各参与机构分别审查	选择性审查	数据标准化方案、跨中心共享协议
商业合作项目	必须审查	选择性审查	数据权属约定、利益冲突管理
涉及遗传/生物标志物数据	必须审查	必须审查	隐私保护措施、长期储存安全性

表 4 医学检验数据分级保护策略

数据级别	示例	最小保护措施	共享范围限制
公开级	检验方法标准化程序	无	无限制
一般敏感级	常规检验群体均值	聚合处理($n>10$)	学术研究机构
高度敏感级	遗传标志物数据、人类免疫缺陷病毒检测结果	匿名化+数据使用协议	签订保密协议的合作方
关键识别级	检验数据与基因组、人脸组合	禁止非必要采集	不可共享

无创产前检测)可能揭示胎儿异常,需禁止在妊娠结局明确前将科研分析结果反馈给临床,避免引发焦虑。(3)知情同意流程优化:使用分级同意书,允许参与者选择不同范围的数据使用授权;为认知障碍者开发可视化知情同意工具(如图文卡片、视频解释)。

推荐意见 22: 建立有效的伦理监督与违规处理机制。(1B, 共识度 96.6%)

解读与证据: 设立独立的监督渠道如伦理委员会、数据保护官,接受和处理与医学检验数据集相关的伦理投诉和举报。制定明确的违规行为认定标准和分级处理程序(如警告、暂停访问、终止合作、追究法律责任等),确保违规行为得到及时、公正、有效的处理,并向相关方和监管机构报告严重违规事件。

《个人信息保护法》要求个人信息处理者建立便捷的个人信息行使权利的受理和处理机制,并规定了对违法行为的处罚措施。有效的监督和惩戒机制是维护伦理规范严肃性的必要保障。透明、公正的处理过程有助于维护信任和制度权威。在此基础上,动态同意可作为传统书面同意的补充机制,依托在线平台允许参与者实时查看数据使用情况并动态调整授权范围,因此尤其适用于诸如糖尿病生物标志物动态监测等需要长期追踪的检验队列研究。平台需提供多语言界面和残障人士辅助功能,确保可及性。

当遭遇创新技术引发的伦理困境(如检验数据训练生成式 AI 模型)时:(1)启动“分阶段审查”:初期限制数据量和应用场;(2)组建“跨学科伦理咨询组”纳入检验医学专家、AI 伦理学家、法律顾问;(3)建立“第三方伦理托管机制”:由中立方暂存数据,待技术风险明确后释放。

二、医学检验数据安全相关

推荐意见 23: 推荐在科研、教学等非临床诊疗场景中,对医学检验数据实行严格的去标识化处理。(1C, 共识度 93.3%)

解读与证据: 在前述隐私优先设计与分级分类

制度框架下,在科研、教学等非临床使用场景中,应对医学检验数据实施严格且系统的去标识化处理,以符合《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》《中华人民共和国数据安全法》及《信息技术安全个人信息去标识化指南》(GB/T 37964—2019)中有关匿名化的要求,确保在依法合规的前提下促进医学数据的安全、高效利用。

去标识化是实现数据可用性与隐私保护平衡的关键环节,除直接标识符(如姓名、身份证号、电话号码、住址等)的删除或替换外,还应防范通过罕见疾病、特殊检验指标组合或极端人口统计学属性(如百岁老人、特殊职业等)导致的再识别风险。常用去标识化技术包括:(1)泛化:将精确值转换为范围或类别,例如将具体年龄转换为年龄段(如 30~39 岁),将详细住址泛化至市或区级行政单位;(2)抑制:删除低频属性或样本量过少的记录,如应用 k-匿名($k\geq 3$)或 L-多样性模型确保任一组合属性的记录数不低于设定阈值;(3)数据合成:采用生成式对抗网络等技术生成符合原始统计分布的人工数据,彻底解除与真实个体的关联^[80]。

处理过程应遵循合法合规原则、安全优先与效用平衡、技术与管理协同、持续监督与改进机制等原则。该过程通常分为 4 个阶段:确定目标、识别标识、处理标识和验证审批。其中,确定目标阶段需明确数据处理场景与共享范围(如完全公开与受限共享的不同风险要求);识别标识阶段应系统识别直接与间接标识符,并依据业务语境分类;处理阶段选择适当的技术模型(如 k-匿名、L-多样性、差分隐私等)与方法;最终需通过内部审批和独立验证方可投入使用。

推荐意见 24: 推荐建立严格的医学检验数据分级授权管理体系。(1B, 共识度 93.3%)

解读与证据: 建立完善的分级授权体系是医院电子病历系统达到高等级的必备条件,在《电子病历系统应用水平分级评价标准(试行)》高等级(5 级及以上)评审要求中,明确提出了对信息系统的数据安全性和权限控制有严格考核。检验数据



访问权限不应“一刀切”,应根据数据敏感性、使用目的及人员角色进行精细化管控。应明确划分数据访问层级,如标识数据、去标识化数据、匿名化数据、聚合统计数据等。《信息安全技术个人信息安全规范》(GB/T 35273—2020)明确规定了个人信息控制者(医院/检验科)应在个人信息处理活动中设置权限控制机制,实现最小必要权限的分配,并对敏感个人信息的访问设置严格的审批流程和日志记录,对所有数据的查询、下载、使用行为进行可追溯的监控^[81]。

推荐意见 25:倡导在安全可控环境下,有序推进检验数据的开放共享与价值释放。(1C, 共识度 93.3%)

解读与证据:数据的价值在于流动和使用,但必须在保障安全、合规和伦理的前提下进行。《国家健康医疗大数据标准、安全和服务管理办法(试行)》中明确规定了健康医疗大数据的采集、存储、应用、共享和开放全流程管理要求。强调在做好脱敏、去标识化的基础上,推动健康医疗大数据在临床、科研、公共卫生等领域的应用,支持在医疗机构内部,为临床诊疗、质量控制和医院管理目的有权限共享^[82]。原则上,对外共享(如科研合作、公共卫生上报)必须使用经过严格去标识化或匿名化处理的数据集。仅限高度聚合的、无法追溯到任何个体患者的统计数据可向公众开放。在跨机构数据协作与分析过程中,应优先采用“数据不出域”的共享模式,利用可信执行环境、联邦学习等隐私计算技术开展协作分析。在联邦学习框架下,各参与机构在本地保留原始医学检验数据,仅在统一任务定义和安全协议下参与模型训练或分析过程,通过共享模型参数或中间统计信息实现多中心协作,从而在不直接交换原始数据的前提下,支持标准数据集在多机构环境中的联合利用与分析。该模式有助于在满足数据安全与合规要求的同时,提升多中心医学检验数据的可用性与协同价值^[83]。若需传输数据,必须通过安全加密通道,并在共享前签订具有法律效力的数据使用协议,明确数据用途、保密责任、销毁期限及违规处罚条款。

此外,在数据共享实践中,应在促进数据利用价值与保障数据安全之间建立动态平衡机制。一方面,通过分级授权、用途限定与审计追踪,确保数据在合法合规前提下被合理使用;另一方面,通过“数据不出域”、联邦学习等技术路径,降低原始数据外流风险。对于共享层级、技术路径及安全强度

的选择,应结合数据敏感性、使用目的与参与机构能力进行分级设定,以在保障隐私与安全的同时,最大化数据集的协同价值。

推荐意见 26:对医学检验数据按影响对象与程度进行分类分级,建立数据持有权制度,衍生数据经统筹授权后进入开放流通体系。(1C, 共识度 93.3%)

解读与证据:参照《数据安全技术数据分类分级规则》(GB/T43697—2024)^[84],医学检验数据为卫生健康数据,基于描述对象分为个人信息(个人身份信息、个人通信信息、个人教育工作信息、医疗机构用户标识符、个人就诊记录、个人基因组测序数据)、检验数据源信息(实验室机构与检验者信息、数据库软件信息)、检验结果数据(样本类型与采集信息、检验项目名称、检测方法学、检验仪器设备信息、检测试剂信息、检测结果、参考区间、危急值标识、检验备注)、检验质量管理数据(检验前/中/周转时间、室内质控数据、室间质评数据)共四类。综合数据的主客体、群体、区域、精度、规模、深度、重要性,及其影响对象与程度,分为核心数据、重要数据、一般数据三个级别。原始数据归其相应主体所有,数据产出方依据服务协议保留客户原始数据的一个存储副本。数据所有方将数据加工形成衍生数据后按《数据资产全过程管理试点方案》形成数据资产,经统筹授权后进入开放流通体系。

需要注意的是,在国家推进数据要素市场化配置和数字中国建设背景下,《数据要素×三年行动计划》明确提出加强数据资源标准化建设和合规流通机制。本共识所倡导的医学检验标准数据集构建框架,在数据结构规范、分类分级管理及安全共享路径等方面,与上述政策导向保持一致,强调在合法合规前提下促进数据要素价值释放。

第六部分 医学检验标准数据集的价值和应用

本章主要阐述医学检验标准数据集作为基础性公共数据资源在临床服务、检验质量管理、科学研究及模型研发评测中的应用价值,明确其在促进医学检验智能化服务规范化、同质化与持续改进中的作用定位。

推荐意见 27:医学检验标准数据集可作为基础性公共数据资源,用于支撑模型研发与性能优化及其典型应用场景落地,并通过统一数据基准支撑模型能力测试与客观评价,从而促进医学检验智能



化服务的规范化、同质化与持续改进。(1B, 共识度 96.6%)

解读与证据:医学检验数据来源多样、方法学复杂,不同机构在项目命名、数据结构、单位及参考区间等方面存在显著差异,若缺乏统一的数据标准与语义规范,将直接制约模型研发、评价和应用。医学检验标准数据集的构建,为检验数据的规范化汇聚与高质量复用提供了基础条件,既可支撑模型研发与性能优化及其典型应用场景的落地,也可作为统一基准支撑模型能力测试与客观评价,从而形成“研发-应用-评价-改进”的良性闭环。

1. 服务临床:常规检验结果解读、检验结果互认状态实时检测与预警、检验组套与检验路径优化。在临床应用中,标准数据集可沉淀“检验项目-人群特征-结果分布-参考区间/决策阈值-临床结局”等关键要素,支撑常规检验结果解读相关的模型训练与知识化表达,提升解释一致性与可迁移性,有助于模型在不同人群和不同机构间的稳定应用。相关应用不仅包括数值型检验结果的智能解读,还可拓展至形态学图像识别、组学检测结果报告辅助及基于检验序列的临床决策支持模型构建等。同时,标准数据集可将检验结果互认所需的项目范围、质量条件及可追溯信息汇聚在数据集中,为基于模型的互认状态监测与异常预警提供支撑。在检验组套与检验路径优化方面,标准数据集通过整合疾病状态、检验项目组合、检测顺序及后续诊疗行为,使不同路径方案具备可分析和可比较的数据基础,从而支撑路径优化相关模型的研发与应用落地^[85]。

2. 服务检验:检验分析前-中-后质量管理、实验室认可认证。在服务检验的应用场景中,若将分析前(采集、运输、标本接收)、分析中(仪器状态、校准、室内质控/室间质评信息)、分析后(复检、危急值、报告更正与解释)等要素纳入统一数据管理,可直接支撑模型研发与性能优化以及典型应用场景落地。在此基础上,可进一步支撑实验室自动化运行监测、报告审核优化以及危急值识别与上报机制等场景。例如,可构建质量监测、异常识别与风险预警模型,推动质量管理从“事后抽查”转向“过程可视与持续改进”。在实验室认可认证方面,相关标准强调质量管理体系的可追溯性与证据完整性^[86]。标准数据集可通过对人员、设备、方法、样本处理及质量控制等要素的统一记录,有助于形成连续、可核验的质量数据基础,支撑认可认证相关应

用的规范化实施。

3. 科学研究:检验项目参考值研究、标志物与检测方法开发与临床研究。在科学研究中,医学检验标准数据集为多中心研究条件下的数据可比性与结果可复现性提供基础支撑^[87]。在检验项目参考值研究中,标准数据集通过统一纳入人群特征、方法学信息及前分析变量,支撑参考值的建立、转移与验证,并提升不同地区/机构间结果解读的一致性。在标志物与检测方法开发方面,标准数据集可统一标志物测量、结局定义与研究设计要素,显著提升标志物发现、验证与临床研究的可复现性,以便于后续向规范化产品与临床转化方向推进。

4. 支撑模型能力测试与客观评价:医学检验标准数据集在支撑模型的开发和调优之外,还承担着统一测试与评价基准的重要功能^[88]。通过对检验项目、结果结构、方法学信息、人群特征及质量要素的标准化表达,标准数据集为不同模型、不同算法版本及不同应用场景提供了可重复、可对比的数据基础,使模型性能评价不再依赖单一机构或特定数据来源^[89]。在此基础上,可对模型在准确性、稳定性及跨机构适用性等方面开展客观测试与比较分析,避免因数据口径不一致导致的评价偏差。同时,统一的数据基准有助于支撑模型在迭代更新、临床试点及质量管理应用中的持续评估,为模型优化、规范应用及风险控制提供可靠依据^[90]。

第七部分 总结与展望

本共识围绕医学检验标准数据集的概念界定、构建原则、核心方法、安全伦理以及应用推广进行了系统梳理,结合当前医学检验数据来源多样、标准不统一和应用需求快速增长的现实问题,凝练形成了具有共识基础和实践导向的推荐意见。

需要明确的是,本共识所提出的医学检验标准数据集构建理念与方法,主要适用于面向 AI 模型开发、验证与评测的标准化数据集建设场景,其应用边界包括但不限于以下几个方面:(1)本共识不等同于医学检验原始业务数据或单一科研课题数据的整理规范,后者可根据具体研究目的采用更为灵活的数据组织方式。(2)本共识所述标准数据集并非直接用于临床决策或模型部署,其主要作用位于模型开发与评估阶段。(3)对于探索性科研数据集、小样本试验性数据或高度定制化研究数据,本共识原则可作为参考,但不构成强制要求。(4)针对



不同层级医疗机构、具体临床场景或检验亚专业的应用策略,需在模型形成并完成验证后,结合实际部署环境另行制定实施方案。通过明确应用边界,有助于避免对标准数据集功能的过度解读,也有利于本共识在实践中得到合理、稳妥的应用。

本共识基于现有临床证据和实践经验,仅针对构建医学检验标准数据集的重要问题形成了推荐意见。部分推荐意见缺乏高质量证据或研究的异质性较高,存在一定的局限性。随着多中心协作研究、真实世界数据应用以及医学检验智能化水平的不断提升,医学检验标准数据集仍需在持续迭代更新、跨机构互认、语义一致性增强及安全合规应用等方面不断完善。同时,应结合前沿技术发展、实际应用反馈以及政策导向变化,动态优化数据要素与质量指标体系,推动标准数据集在科研、临床和管理等多场景中的深度应用,持续提升医学检验智能化发展的科学性、可比性与规范性。

本共识制订专家委员会名单

牵头专家:尚红[中国医科大学附属第一医院检验科 医疗辅助技术(医学检验)国家临床医学研究中心];陈瑜(浙江大学医学院附属第一医院检验科);韩晓旭[中国医科大学附属第一医院检验科 医疗辅助技术(医学检验)国家临床医学研究中心]

共识执笔专家(按姓氏汉语拼音排序):杜忠礼(国家卫生健康委临床检验中心);段昕岑(复旦大学附属中山医院检验科);费春荣(浙江大学医学院附属邵逸夫医院检验科);何彦(南方医科大学珠江医院检验医学部);贺娟(内蒙古自治区人民医院临床检验医学中心);蒋文斌(贵州省人民医院检验科);李娟(山东大学齐鲁第二医院检验医学中心);李芮冰(解放军总医院第一医学中心检验科);李世奇(福建医科大学附属第一医院检验科);梁运来(中南大学湘雅医院检验科);刘灿(福建医科大学附属第一医院检验科);鲁大鹏(中国科学技术大学附属第一医院检验科);孟芝君(山西省人民医院检验科);苗强(四川大学华西医院实验医学科);穆原(南京医科大学第一附属医院检验学部);潘俊希(昆明医科大学附属第一医院暨云南省医学检验临床医学研究中心);任章银(陆军军医大学第一附属医院检验科);王连明(哈尔滨医科大学附属第一医院检验科);王文博(华中科技大学同济医学院附属同济医院检验科);吴丹(浙江大学医学院附属邵逸夫医院检验科);肖盟(中国医学科学院北京协和医院检验科);徐凯(海南省人民医院检验科);杨汝(宁夏医科大学总医院医学实验中心);袁旭(华中科技大学同济医学院附属同济医院检验科);俞秋阳(吉林大学第一医院检验科);章海斌(南昌大学第二附属医院检验科);张瑞(河北医科大学第二医院检验科);张涛(空军军医大学第一附属医院检验科);张兴旺(甘肃省人民医院检验

科);赵俊伟(郑州大学第一附属医院医学检验科);郑辉(中国科学技术大学附属第一医院检验科);郑书发(浙江大学医学院附属第一医院检验科);周春雷(天津市第一中心医院检验科)

共识制订专家(按姓氏汉语拼音排序):阿祥仁(青海省人民医院 青海省临床检验中心);陈捷(四川大学华西医院实验医学科);陈鸣(陆军军医大学第一附属医院检验科);陈文祥(国家卫生健康委临床检验中心);陈瑜(浙江大学医学院附属第一医院检验科);程黎明(华中科技大学同济医学院附属同济医院检验科);崔丽艳(北京大学第三医院检验科);崔巍(中国医学科学院肿瘤医院检验科);单斌(昆明医科大学附属第一医院暨云南省医学检验临床医学研究中心);丁海涛(内蒙古自治区人民医院临床检验医学中心);段勇(昆明医科大学附属第一医院暨云南省检验医学重点实验室);高小玲(海南省人民医院检验科);关洁(哈尔滨医科大学附属第一医院检验科);关明(复旦大学附属华山医院检验医学科);郭玮(复旦大学附属中山医院检验科);韩晓旭[中国医科大学附属第一医院检验科 医疗辅助技术(医学检验)国家临床医学研究中心];吉建民(山西省人民医院检验科);李锋(宁夏医科大学总医院医学实验中心);李辉(新疆医科大学第一附属医院医学检验中心);李莉(上海临床研究中心);李绵洋(解放军总医院第一医学中心检验科);梁红萍(山西省人民医院检验科);刘治娟(西藏自治区人民医院检验科);姜加陶(上海市第一人民医院检验医学中心);马洁(国家卫生健康委临床检验中心);毛海婷(山东大学齐鲁第二医院检验医学中心);穆红(天津市第一中心医院检验科);欧启水(福建医科大学附属第一医院检验科);潘柏中(复旦大学附属中山医院检验科);潘世扬(南京医科大学第一附属医院检验学部);尚红[中国医科大学附属第一医院检验科 医疗辅助技术(医学检验)国家临床医学研究中心];史敏(河北医科大学第二医院检验科);魏莲花(甘肃省人民医院检验科);王蓓丽(复旦大学附属中山医院检验科);王保龙(中国科学技术大学附属第一医院检验科);王传新(山东大学齐鲁第二医院检验医学中心);王健(广西医科大学第一附属医院检验科);王前(南方医科大学珠江医院检验医学部);王万海(郑州大学第一附属医院医学检验科);王小中(南昌大学第二附属医院检验科);王一多(哈尔滨医科大学附属第一医院检验科);谢鑫友(浙江大学医学院附属邵逸夫医院检验科);许建成(吉林大学第一医院检验科);杨大千(浙江大学医学院附属第一医院检验科);杨启文(中国医学科学院北京协和医院检验科);易斌(中南大学湘雅医院检验科);应斌武(四川大学华西医院实验医学科);张华(贵州省人民医院检验科);张钧(浙江大学医学院附属邵逸夫医院检验科);周磊(空军军医大学第一附属医院检验科)

共识执笔统筹专家(按姓氏汉语拼音排序):马德龙[医疗辅助技术(医学检验)国家临床医学研究中心];杨晓陶[医疗辅助技术(医学检验)国家临床医学研究中心];张歆[医疗



辅助技术(医学检验)国家临床医学研究中心]

利益冲突 所有作者声明不存在利益冲突

参 考 文 献

- [1] Sikaris KA. Enhancing the clinical value of medical laboratory testing[J]. Clin Biochem Rev, 2017, 38(3): 107-114.
- [2] 国务院. 国务院关于印发新一代人工智能发展规划的通知[EB/OL]. (2017-07-20) [2025-12-20]. <https://www.nhc.gov.cn/bgt/gwywj2/201707/3b2a93a71c794d9c8137ab394b21d8f3.shtml>.
- [3] 国家数据局, 中央网信办, 科技部, 等. 十七部门关于印发《“数据要素×”三年行动计划(2024—2026 年)》的通知[EB/OL]. (2023-12-31) [2025-11-10]. https://www.cac.gov.cn/2024-01/05/c_1706119078060945.htm.
- [4] 国家数据局综合司. 数字中国建设 2025 年行动方案[EB/OL]. (2025-05-16) [2025-11-10]. https://www.gov.cn/lianbo/bumen/202505/content_7024041.htm.
- [5] Mandel JC, Kreda DA, Mandl KD, et al. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records[J]. J Am Med Inform Assoc, 2016, 23(5):899-908. DOI: 10.1093/jamia/ocv189.
- [6] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence[J]. Nat Med, 2019, 25(1): 44-56. DOI: 10.1038/s41591-018-0300-7.
- [7] Xu X, Zhang Y, Wang Y, et al. Impact of regional digital transformation on public health: an empirical analysis based on 31 provinces in China[J]. BMC Public Health, 2025, 25(1):2485. DOI: 10.1186/s12889-025-23670-8.
- [8] Kleinert P, Gebhardt M, Buckow K, et al. Germany's approach for the secondary use of federated real-world data in the German portal for medical research data[J]. Stud Health Technol Inform, 2024, 316:1704-1708. DOI: 10.3233/SHIT240751.
- [9] Hahn W, Ahmadi N, Hoffmann K, et al. Synthetic data generation in hematology-paving the way for OMOP and FHIR integration[J]. Stud Health Technol Inform, 2024, 316:1472-1476. DOI: 10.3233/SHIT240692.
- [10] Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records[J]. NPJ Digit Med, 2018, 1:18. DOI: 10.1038/s41746-018-0029-1.
- [11] Seoni S, Shahini A, Meiburger KM, et al. All you need is data preparation: a systematic review of image harmonization techniques in Multi-center/device studies for medical support systems[J]. Comput Methods Programs Biomed, 2024, 250: 108200. DOI: 10.1016/j.cmpb.2024.108200.
- [12] 王浩, 孟祥峰, 郝烨, 等. 行业标准《人工智能医疗器械质量要求和评价第 2 部分: 数据集通用要求》解析[J]. 协和医学杂志, 2023, 14(6): 1180-1184. DOI: 10.12290/xhyxzz.2023-0464.
- [13] Luo J, Wu M, Gopukumar D, et al. Big data application in biomedical research and health care: a literature review[J]. Biomed Inform Insights, 2016, 8:1-10. DOI: 10.4137/BII.S31559.
- [14] Turner K, Hohman KH. Demonstrated progress and future promise of chronic disease data modernization[J]. Prev Chronic Dis, 2024, 21:E86. DOI: 10.5888/pcd21.240396.
- [15] Huang SC, Pareek A, Seyyedi S, et al. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines[J]. NPJ Digit Med, 2020, 3:136. DOI: 10.1038/s41746-020-00341-z.
- [16] Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future[J]. Stroke Vasc Neurol, 2017, 2(4): 230-243. DOI: 10.1136/svn-2017-000101.
- [17] Cui C, Yang H, Wang Y, et al. Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: a review[J]. Prog Biomed Eng (Bristol), 2023, 5(2): 10.1088/2516-1091/acc2fe. DOI: 10.1088/2516-1091/acc2fe.
- [18] Delgado Rodríguez JA, Pastor García MI, Gómez Cobo C, et al. Assessment of a laboratory critical risk result notification protocol in a tertiary care hospital and their use in clinical decision making[J]. Biochem Med (Zagreb), 2019, 29(3):030703. DOI: 10.11613/BM.2019.030703.
- [19] Maleki Varnosfaderani S, Forouzanfar M. The role of AI in hospitals and clinics: transforming healthcare in the 21st century[J]. Bioengineering (Basel), 2024, 11(4):337. DOI: 10.3390/bioengineering11040337.
- [20] 国家市场监督管理总局, 中国国家标准化管理委员会. 信息安全技术-健康医疗数据安全指南: GB/T 39725—2020[S]. 北京: 国家标准委, 2020.
- [21] Price WN, Cohen IG. Privacy in the age of medical big data[J]. Nat Med, 2019, 25(1): 37-43. DOI: 10.1038/s41591-018-0272-7.
- [22] Feng J, Xia F, Singh K, et al. Not all clinical AI monitoring systems are created equal: review and recommendations[J]. Nejm Ai, 2025, 2(2). DOI: 10.1056/AIra2400657.
- [23] 中华医学会检验医学分会, 中国医师协会急诊医师分会, 全军急救医学专业委员会. 急诊检验能力建设与规范中国专家共识[J]. 解放军医学杂志, 2020, 45(1):21-42. DOI: 10.11855/j.issn.0577-7402.2020.01.02.
- [24] Clinical and Laboratory Standards Institute. Establishing molecular testing in medical laboratory environments: MM19, 2nd ed[S]. Wayne, PA: CLSI, 2025.
- [25] 国家卫生健康标准委员会卫生健康信息标准专业委员会, 国家卫生健康委统计信息中心. 国家医疗健康信息医院信息互联互通标准化成熟度测评方案(2020 年版)[EB/OL]. (2020-08-06) [2025-11-12]. https://www.nhc.gov.cn/mohwsbwstjxxzx/s8553/202008/bdd0d4fcb1c747dda00e0adec3c17b9/files/1740035892364_40918.pdf.
- [26] Çubukçu HC, Boursier G, Linko S, et al. Regulating the future of laboratory medicine: European regulatory landscape of AI-driven medical device software in laboratory medicine[J]. Clin Chem Lab Med, 2025, 63(10): 1891-1914. DOI: 10.1515/cclm-2025-0482.
- [27] World Health Organization. Laboratory quality management system: handbook[EB/OL]. [2025-11-15]. Geneva: World Health Organization, 2011.
- [28] The Regenstrief Institute, UCUM Editorial Board. The UCUM specification[EB/OL]. [2025-11-15]. <https://ucum.org/ucum.html>.
- [29] U.S. Food and Drug Administration. Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan[EB/OL]. (2021-01) [2025-11-20]. <https://www.fda.gov/media/145022/download>.
- [30] World Health Organization. Global strategy on digital health 2020—2025[EB/OL]. Geneva: World Health



- Organization, 2021[2025-11-20]. <https://www.who.int/publications/i/item/9789240020924>.
- [31] Shadbahr T, Roberts M, Stanczuk J, et al. The impact of imputation quality on machine learning classifiers for datasets with missing values[J]. *Commun Med (Lond)*, 2023, 3(1):139. DOI: 10.1038/s43856-023-00356-z.
- [32] National Institute for Health and Care Excellence (NICE). Medical technologies evaluation programme methods guide[EB/OL]. (2017-08-21) [2025-11-20]. <https://www.nice.org.uk/process/pmg33/>.
- [33] JCGM, BIPM. JCGM 100: 2008 guide to the expression of uncertainty in measurement[EB/OL]. Sèvres, France: BIPM, 2008[2025-11-22]. https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf.
- [34] 常冰洁, 赵琳琳, 吴浩, 等. 数字医疗赋能全科医学: 转型路径、挑战与应对策略[J]. *中华全科医师杂志*, 2025, 24(8): 916-924. DOI: 10.3760/cma.j.cn114798-20250423-00348.
- [35] Strasberg HR, Rhodes B, Del Fiore G, et al. Contemporary clinical decision support standards using Health Level Seven International Fast Healthcare Interoperability Resources[J]. *J Am Med Inform Assoc*, 2021, 28(8): 1796-1806. DOI: 10.1093/jamia/ocab070.
- [36] ASTM International. Standard specification for low-level protocol to transfer messages between clinical laboratory instruments and computer systems (withdrawn 2002): ASTM E1381-02[S]. West Conshohocken, PA: ASTM International, 2002: 1-11. <https://www.astm.org/e1381-02.html>.
- [37] Adamusiak T, Shimoyama N, Shimoyama M. Next generation phenotyping using the unified medical language system[J]. *JMIR Med Inform*, 2014, 2(1):e5. DOI: 10.2196/medinform.3172.
- [38] Williams A, Mekhail AM, Williams J, et al. Effective resource management using machine learning in medicine: an applied example[J]. *BMJ Simul Technol Enhanc Learn*, 2019, 5(2): 85-90. DOI: 10.1136/bmjstel-2017-000289.
- [39] Vorisek CN, Lehne M, Klopfenstein S, et al. Fast healthcare interoperability resources (FHIR) for interoperability in health research: systematic review[J]. *JMIR Med Inform*, 2022, 10(7):e35724. DOI: 10.2196/35724.
- [40] Wang KC, Patel JB, Vyas B, et al. Use of radiology procedure codes in health care: the need for standardization and structure[J]. *Radiographics*, 2017, 37(4):1099-1110. DOI: 10.1148/rg.2017160188.
- [41] All of Us Research Program Investigators, Denny JC, Rutter JL, et al. The "All of Us" research program[J]. *N Engl J Med*, 2019, 381(7): 668-676. DOI: 10.1056/NEJMs1809937.
- [42] Allen NE, Lacey B, Lawlor DA, et al. Prospective study design and data analysis in UK biobank[J]. *Sci Transl Med*, 2024, 16(729): ead4428. DOI: 10.1126/scitranslmed.adf4428.
- [43] International Organization for Standardization. ISO 15189: 2022 medical laboratories-requirements for quality and competence[S]. Geneva: ISO, 2022.
- [44] 刘畅, 彭晓霞, 蔡思雨, 等. 多中心临床实验室来源的真实世界数据预处理流程构建[J]. *中华流行病学杂志*, 2025, 46(2): 296-306. DOI: 10.3760/cma.j.cn112338-20240620-00364.
- [45] 曲琳琳, 续薇. 聚焦 AI 血液病预测模型 关注模型构建与效能[J]. *中华检验医学杂志*, 2025, 48(5): 543-548. DOI: 10.3760/cma.j.cn114452-20250217-00086.
- [46] 张路, 王薇, 何法霖, 等. 临床检验检验前质量指标的一致化[J]. *现代检验医学杂志*, 2016 (1): 158-160, 164. DOI: 10.3969/j.issn.1671-7414.2016.01.047.
- [47] 王静, 金霞霞, 卢国光, 等. 有核红细胞联合 Ranson 评分及 APACHE II 评分构建重症急性胰腺炎患者的结局预测模型[J]. *中华检验医学杂志*, 2020, 43(1):63-70. DOI: 10.3760/cma.j.issn.1009-8158.2020.01.007.
- [48] 王业青, 时代, 陆宽, 等. 预测脑出血血肿扩大的诺模图模型建立与多角度评价[J]. *中华医学杂志*, 2021, 101(31): 2471-2477. DOI: 10.3760/cma.j.cn112137-20210118-00161.
- [49] Suzuki G, Nishioka S, Kobori T, et al. SACrA score to predict the initiation of renal replacement therapy in critically ill patients: a single-center retrospective study[J]. *Ren Fail*, 2024, 46(2): 2404237. DOI: 10.1080/0886022X.2024.2404237.
- [50] 谢政, 金晶, 刘东松, 等. 基于监督机器学习算法构建脓毒性休克患者死亡风险的预测模型[J]. *中华危重病急救医学*, 2024, 36(4): 345-352. DOI: 10.3760/cma.j.cn121430-20230930-00832.
- [51] Stelzer EM, Höltege J, Zhou N, et al. Cross-cultural generalizability of the ICD-11 PGD symptom network: identification of central symptoms and culturally specific items across German-speaking and Chinese bereaved[J]. *Compr Psychiatry*, 2020, 103: 152211. DOI: 10.1016/j.comppsy.2020.152211.
- [52] 何丹毅, 樊泽, 吴宇飞, 等. 加速康复外科临床路径对正颌手术患者结局的影响: 一项回顾性研究[J]. *空军军医大学学报*, 2023, 44(9): 819-825. DOI: 10.13276/j.issn.2097-1656.2023.09.005.
- [53] Wei X, Chen X, Zhang Z, et al. Risk analysis of the association between different hemoglobin glycation index and poor prognosis in critical patients with coronary heart disease-a study based on the MIMIC-IV database[J]. *Cardiovasc Diabetol*, 2024, 23(1): 113. DOI: 10.1186/s12933-024-02206-1.
- [54] Peng S, Huang J, Liu X, et al. Interpretable machine learning for 28-day all-cause in-hospital mortality prediction in critically ill patients with heart failure combined with hypertension: a retrospective cohort study based on medical information mart for intensive care database-IV and eICU databases[J]. *Front Cardiovasc Med*, 2022, 9:994359. DOI: 10.3389/fcvm.2022.994359.
- [55] Li J, Guo S, Ma R, et al. Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets[J]. *BMC Med Res Methodol*, 2024, 24(1):41. DOI: 10.1186/s12874-024-02173-x.
- [56] Sun W, Zhang L, Mou D, et al. Interpretable machine learning model for early mortality prediction in septic patients using routine post-diagnosis clinical data: a multicenter study[J]. *J Inflamm Res*, 2025, 18: 15003-15015. DOI: 10.2147/JIR.S553042.
- [57] 全国数据标准化技术委员会. 高质量数据集 分类指南: TC609-5-2025-03[S]. 北京: 全国数据标准化技术委员会, 2025.
- [58] 国家数据局. 高质量数据集建设指引[EB/OL]. (2025-08-28) [2025-11-22]. <https://www.nda.gov.cn/sjj/ywpl/szkjyjc/0830/ff808081-96b465bf-0198-fb119f46-16b1.pdf>.
- [59] 国家药品监督管理局. 人工智能医疗器械质量要求和评价 第 2 部分: 数据集通用要求: YY/T 1833.2—2022[S]. 北京:



- 中国标准出版社, 2022.
- [60] 全国数据标准化技术委员会. 高质量数据集 质量评测规范: TC609-5-2025-04[S]. 北京: 全国数据标准化技术委员会, 2025.
- [61] 国家卫生健康标准委员会卫生健康信息标准专业委员会. 卫生健康信息元目录. 第 9 部分: 实验室检查: WS/T 363.9—2023[S]. 北京: 国家卫生健康委统计信息中心, 2023.
- [62] ISO. Medical laboratories--requirements for quality and competence: ISO 15189 : 2022[S]. Geneva: International Organization for Standardization, 2022.
- [63] Lin LI. A concordance correlation coefficient to evaluate reproducibility[J]. *Biometrics*, 1989, 45(1):255-268.
- [64] 中国通信标准化协会. 医疗健康大数据平台 质量管理技术要求和测试方法: YD/T 4900-2024[S]. 北京: 工业和信息化部, 2024.
- [65] 卫生部信息化工作领导小组办公室. 卫生综合管理信息平台建设指南(试行)[EB/OL]. (2011-04-28) [2025-11-22]. <https://www.nhc.gov.cn/cmsresources/mohwsbwstjxxxz/cmsrsdocument/doc11829.pdf>.
- [66] Blatter TU, Witte H, Nakas CT, et al. Big data in laboratory medicine-FAIR quality for AI? [J]. *Diagnostics (Basel)*, 2022, 12(8):1923. DOI: 10.3390/diagnostics12081923.
- [67] Health Level Seven International. HL7 FHIR release 4: specification[EB/OL]. (2019-10-30) [2025-08-21]. <https://www.hl7.org/FHIR/R4/>.
- [68] Office of the National Coordinator for Health IT. United States Core Data for Interoperability (USCDI), Version 3[R]. Washington, DC: ONC, 2022.
- [69] LOINC Committee. LOINC version 2.81 release highlights [EB/OL]. Regenstrief Institute, 2025[2025-11-22]. <https://loinc.org/news/loinc-version-2-81-release-highlights/>.
- [70] European Commission. Regulation (EU) 2017/746 of the European Parliament and of the Council on in vitro diagnostic medical devices. Official Journal of the European Union, 2017[EB/OL]. [2025-11-22]. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32017R0746&from=EN>.
- [71] 中华人民共和国国家卫生健康委员会. 电子病历共享文档规范: WST 500[S]. 北京: 国家标准化管理委员会, 2016.
- [72] 中华人民共和国国家卫生健康委员会. 国家健康医疗大数据标准、安全和服务管理办法(试行)[S]. 北京: 国家卫生健康委员会, 2018.
- [73] Plebani M. Laboratory medicine: value for patients is the goal[J]. *Clin Chem*, 2007, 53(10): 1873-1874. DOI: 10.1373/clinchem.2007.090563.
- [74] Topalian SL, Taube JM, Anders RA, et al. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy[J]. *Nat Rev Cancer*, 2016, 16(5):275-287. DOI: 10.1038/nrc.2016.36.
- [75] U. S. Food and Drug Administration. In vitro diagnostics EUAs[EB/OL]. Silver Spring, MD: FDA, 2023[2025-11-25]. <https://www.fda.gov/medical-devices/COVID-19-emergency-use-authorizations-medical-devices/in-vitro-diagnostics-euas>.
- [76] 国家药品监督管理局. 关于发布真实世界数据用于医疗器械临床评价的技术指导原则(试行)的通告(2020 年第 77 号)[EB/OL]. 北京: 国家药品监督管理局(2020-11-24) [2025-11-25]. <https://www.nmpa.gov.cn/xxgk/ggtg/ylqxggtg/ylqxqtggtg/20201126090030150.html>.
- [77] Plebani M. Harmonization in laboratory medicine: the complete picture[J]. *Clin Chem Lab Med*, 2013, 51(4): 741-751. DOI: 10.1515/cclm-2013-0075.
- [78] Ceriotti F, Cobbaert C. Harmonization of external quality assessment schemes and their role-clinical chemistry and beyond[J]. *Clin Chem Lab Med*, 2018, 56(10):1587-1590. DOI: 10.1515/cclm-2018-0265.
- [79] World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects[J]. *JAMA*, 2013, 310(20):2191-2194. DOI: 10.1001/jama.2013.281053.
- [80] Yoon J, Mizrahi M, Ghalaty NF, et al. EHR-Safe: generating high-fidelity and privacy-preserving synthetic electronic health records[J]. *NPJ Digit Med*, 2023, 6(1): 141. DOI: 10.1038/s41746-023-00888-7.
- [81] 全国网络安全标准化技术委员会. 信息安全技术 个人信息安全规范: GB/T 35273-2020[S]. 北京: 中国标准出版社, 2020.
- [82] 国家卫生健康委员会. 关于印发国家健康医疗大数据标准、安全和服务管理办法(试行)的通知[EB/OL]. (2018-09-13) [2025-11-26]. <https://www.nhc.gov.cn/guihuaxxs/c100133/201809/c5e3e58313a24fdeb26d3dc07003d56e.shtml>.
- [83] Lekadir K, Frangi AF, Porras AR, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare[J]. *BMJ*, 2025, 388:e081554. DOI: 10.1136/bmj-2024-081554.
- [84] 单博深, 左晓栋. 国标《数据安全技术 数据分类分级规则》分析与解读[J]. *网络安全与数据治理*, 2024, 43(11):19-22. DOI: 10.19358/j.issn.2097-1788.2024.11.004.
- [85] Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence-what is it and what can it tell us? [J]. *N Engl J Med*, 2016, 375(23): 2293-2297. DOI: 10.1056/NEJmsb1609216.
- [86] 中国合格评定国家认可委员会. 医学实验室质量和能力认可准则: CNAS-CL02: 2023[S]. 北京: 中国合格评定国家认可委员会, 2023.
- [87] Sheehan J, Hirschfeld S, Foster E, et al. Improving the value of clinical research through the use of common data elements[J]. *Clin Trials*, 2016, 13(6): 671-676. DOI: 10.1177/1740774516653238.
- [88] McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening[J]. *Nature*, 2020, 577(7788): 89-94. DOI: 10.1038/s41586-019-1799-6.
- [89] Karargyris A, Umeton R, Sheller MJ, et al. Federated benchmarking of medical artificial intelligence with MedPerf[J]. *Nat Mach Intell*, 2023, 5(7): 799-810. DOI: 10.1038/s42256-023-00652-2.
- [90] Wells BJ, Nguyen HM, McWilliams A, et al. A practical framework for appropriate implementation and review of artificial intelligence (FAIR-AI) in healthcare[J]. *NPJ Digit Med*, 2025, 8(1):514. DOI: 10.1038/s41746-025-01900-y.

