



首都医科大学附属北京友谊医院信息中心主任

吕晗:从“数据大”到“大数据”是新时代命题

因此,建立健全并严格执行数据质量控制体系,确保数据的真实性、准确性、完整性、一致性和及时性,已成为保障临床科研成果可靠性、推动研究结论有效转化的核心前提与根本保障。

而一旦拥有了临床或科研的高质量数据集,所带来的收益将是多维度且深远的。在临床实践层面,高质量数据集能够为精准医疗的实现奠定坚实基础。通过对大量标准化、结构化的患者临床数据(如病史、症状、体征、实验室检查结果、影像学资料、治疗方案及预后情况等)进行深度分析,可帮助医

生更准确地进行疾病特征的挖掘,助力临床实现对疾病的发生、进展及并发症风险的早期预警,改善患者预后并合理配置医疗资源。

在科研创新领域,高质量数据集是推动医学科学突破的核心驱动力。它为研究人员提供了前所未有的机会去探索疾病的复杂机制,发现新的疾病标志物、药物靶点 and 治疗方法。通过对大规模纵向数据的分析,可以揭示疾病的自然病程和演变规律;而多中心、跨区域的数据集则有助于验证研究成果的普适性和可靠性,加速科研

成果向临床应用的转化。

此外,高质量数据集还能促进基础研究与临床研究的深度融合,使基础研究的发现能够更快地在临床场景中得到检验和应用。反之,临床中遇到的问题也能更精准地反馈给基础研究,形成“从临床到基础,再从基础到临床”的良性循环,极大地缩短新药研发和新疗法探索的周期。

我们要牢牢守住数据质量这一生命线,让“大数据”真正成为驱动临床科研创新、提升医疗服务水平的强大引擎。

当前,临床科研的核心矛盾已发生根本性转变,不再是“有没有数”的问题,而是聚焦于“数能不能用、敢不敢用”这一关键命题。因此,从“数据大”到“大数据”的演进,标志着临床科研领域已跨越了数据稀缺的初级阶段。

在这一背景下,数据质量控制环节的重要性被提升到前所未有的高度。



中国科学院生物物理研究所研究员

李岩:数据可持续性与共享性是关键议题

性的人群队列,为中国特色慢性病研究奠定了坚实基础。诸多的国内队列填补了中国人群慢性病病因空白,以本土化证据支撑政策与指南更新,推动我国精准医学迈向国际前沿。

当前研究范式正从传统暴露因素分析向多组学整合方向深化。传统流行病学调查结合高通量组学技术的新模式,为揭示疾病发生发展的分子机制提供了全新视角。多个大型队列项目均在开展大规模全基因组测序工

作,旨在构建中国人群特有的遗传变异图谱;中国慢性病前瞻性研究项目、中国自然人群生物资源库等项目也在积极推进多维组学云平台建设,致力于从基因组、蛋白质组等多层次解析复杂疾病的生物学基础。这种多组学整合策略不仅有助于发现新的生物标志物,还能深入揭示基因-环境交互作用的复杂网络。

在数字化时代到来时,确保数据的可持续性和共享性是大型队列未来发展

的关键议题。在严格遵守伦理规范和隐私保护的前提下,建立高效、安全的数据共享机制,将极大提升队列研究的科学价值和社会效益。

展望未来,大型人群队列研究将继续在慢性病防控、健康老龄化、精准医学等领域发挥重要作用。通过多学科协作、技术创新和资源整合,这些研究将为完善卫生政策、优化临床实践提供更加有力的科学依据,最终为实现全民健康目标作出贡献。

在全球范围内,大型人群队列研究已形成相当规模。英国生物样本库、欧洲癌症与营养前瞻性调查等国际知名项目积累了丰富经验。我国的大型人群队列项目虽起步相对较晚,但通过严谨的设计和高质量的随访,已建立起具有代表



北京协和医院临床生物样本中心主任

郭丹:人工智能时代医疗机构样本库将呈现“四化”

析临床信息、采样记录、转运日志等“接踵而来”的数据流内容,抓取其中的术语偏差、数值冲突、流程缺项等,实时生成纠偏提示与质量评分,把不规范、不完整、无法溯源的样本信息拦截在入库前,实现样本数据标准化管理的“实时化”。

其二,针对多模态、跨尺度数据的整合,实现数据孪生化。样本库通过整合影像、病理、多组学数据等,借助人工智能技术可以把这些多模态的数据碎片拼成患者的“数字孪生体”。在肿瘤多模态诊疗和预后评估、药物设计与筛选等方向,研究者可进一步依托算力和大语言模型等,从云端验证模型到真实世界落地,降低试错成本,提高验证效率。

其三,针对样本和数据的共享机制,实现共享智能化。基于隐私计算、联邦学习等技术体系下的智能共享网络有望实现“数据不出院、模型多院跑”的应用场景,通过区块链技术确保生物样本数据的真实性和可追溯性,保护数据隐私,提高共享效率与透明度,协同促进样本库的同质化管理和资源的共享应用。

其四,针对样本和数据的转化利用,实现样本库发展的可持续化。样本库依托高质量生物样本资源、标准化数据集与数字化模型等,通过规范化的样本共享与数据交易平台助力资源的确权和合规流通,实现患者获益、科研共享、产业转化的可持续生态,并进一步通过数据

可信区间的建立,打造医学研究+产业应用共创模式,构建样本和数据资源高效流通机制,依托转化应用及产业化生态体系,形成长效发展模式。

以 AI 淬炼品质,让每份样本成为可信的“种子”;以孪生拓展边界,把真实世界升维为无限试验场;以共享疏通脉络,使数据在流动中增殖;以转化提升价值,让科研、临床与产业在此交汇共赢。医疗机构样本库将在人工智能时代下快速发展,焕发无限活力,通过“四化”成为连通基础研究、临床与转化研究的重要枢纽,以及预防、诊疗、康复、健康管理等全链条的强力支撑,驱动精准医疗时代的新质生产力。

医疗机构样本库作为医药卫生和科技健康领域创新的重要基础性支撑,未来将在人工智能赋能下,依托高质量数据集和可信数据空间+大规模算力与模型,在标准化和信息化的基础上,呈现“四化”的趋势,以快速拥抱智能化时代的到来。

其一,针对多元化样本和数据全流程管理,做到质量过程管理实时化。传统质量管理模式依靠人工抽查,依赖于事后检查,无法实时纠错。AI 可以辅助解