

# 健康大数据在公共卫生领域中的应用与挑战

马瑞雪

武城县爱国卫生运动推进中心, 山东省德州市, 253300;

**摘要:** 健康大数据把医疗机构诊疗记录、公共卫生监测信息和居民健康档案等多种不同的数据融合起来, 给公共卫生管理与决策赋予了新的角度和办法。本文系统地论述了健康大数据在疾病监测预警、流行病学研究、卫生资源优化配置、健康风险评估等各个方面的应用模式和价值, 并对数据采集、整合分析、跨部门共享等环节遇到的关键问题进行了深入的研究。就目前存在的数据质量参差不齐、隐私安全风险、标准化程度不高、专业人才缺乏等主要问题, 提出创建完善的数据治理体系、加强数据安全和隐私保护、创建高效的数据共享平台、培养复合型专业人才等对策。本文意在促进健康大数据在公共卫生领域深入运用, 提高公共卫生服务整体效率提供参考和实践指导。

**关键词:** 健康大数据; 公共卫生; 疾病监测; 数据治理

**DOI:** 10.69979/3041-0673.26.04.079

公共卫生工作包含疾病防控、健康促进、应急响应等各个方面, 传统依靠抽样调查、被动报告的方式获取信息存在覆盖面小、时效性差等缺点。健康大数据技术的发展使得大量的、多种类的、实时的健康信息得以聚集, 并且可以进行深入的分析, 从而实现了人群健康风险的精准识别, 资源配置的科学优化, 应对突发公共卫生事件的有效性。挖掘健康大数据的价值之后, 公共卫生决策模式由原来的依靠经验转变为依靠数据, 公共卫生服务也朝着更加精准化、智能化、个性化的方向发展, 为提高全人群健康水平、建设强大的公共卫生体系开辟了新的途径。

## 1 健康大数据在疾病监测预警中的应用

### 1.1 传染病智能监测系统

健康大数据技术给传染病监测提供多种信息。采用自然语言处理技术提取电子病历的症状特征, 可以起到辅助诊断疑似病例的作用。实验室检测结果实时上报同临床信息融合起来, 改善了病例认定的准确性。监测药店某种药品的销量变化, 可以发现潜在疫情的迹象。通过分析搜索引擎查询以及社交媒体上的内容, 可以感知公众对于某种症状或者疾病关注度的变化。多源信息经过数据融合算法整合以后可以形成更灵敏的预警指标体系。使用时序分析技术, 系统可以自动找到发病率异常波动, 当其超过设定阈值之后便立刻向卫生部门发出预警信息, 从而把公共卫生事件的应对从被动的响应变为主动的预警。

### 1.2 慢性病管理与风险预测

对于慢性病的长期防控, 健康大数据支撑创建个体化的风险评价和管理机制。整合居民电子健康档案、体检数据、可穿戴设备记录、医保信息等各方面的数据, 形成动态更新的个人健康全貌。基于机器学习算法, 可以对年龄、生活方式、生化指标等各方面风险因素进行综合分析, 预测一个人将来得病的概率<sup>[1]</sup>。系统可以自动筛选出高风险人群, 给高风险人群推送有针对性的健康提醒和干预建议。在此基础上实行分级管理, 对高风险者加强定期随访、监测, 对普通人群侧重健康教育、行为指导, 实现资源的优化配置、健康效益最大化。

### 1.3 突发公共卫生事件应急响应

突发公共卫生事件当中, 健康大数据技术属于开展快速、精准应急响应的重要依靠。使用统一病例信息管理平台对疫情时空动态做实时汇总、分析、可视化展示。数字化流调工具可以对患者的轨迹、接触史进行高效的采集, 并且还可以和移动通信等数据结合使用, 从而达到精准追踪传播链、判断扩散风险的目的。数学模型可以模拟出不同的防控情景下疫情的发展趋势, 给科学决策提供量化的依据。资源调度平台可以集成医疗物资、人员、设施等各方面数据, 使应急资源可以得到快速调配。依靠信息发布系统及时、准确地传递疫情信息及防护指南, 以指导社会面对疫情采取正确的应对措施, 从而提高防控整体效能。

## 2 健康大数据支持的流行病学研究

### 2.1 疾病危险因素识别

健康大数据极大地加深并拓宽了流行病学研究。基

于电子健康档案的队列研究可以获得长时间段内连续的个体数据,减少传统研究中由于失访或者信息偏倚造成的误差,进而更加准确地分析暴露因素与疾病结局之间的联系。病例对照研究可以利用大样本的医疗数据库迅速筛选研究对象。生态学研究把地理信息系统同起来,探究区域疾病分布同环境、社会经济因素的空间联系。多组学数据的整合分析(基因组、代谢组等)可以从小分子角度出发去探究疾病机制。采用因果推断等高级统计方法,可以较好地处理观察性数据中的混杂因素,给危险因素的评价提供更加可靠的因果证据。

## 2.2 疾病负担评估与预测

健康大数据给精准评估和前瞻预测疾病负担提供强大的支撑。综合利用医疗服务量、医保支付等数据,可以对疾病的消耗造成的直接医疗资源和费用进行系统的测算。结合患者报告结局等信息可以更全面的量化疾病对生命质量的影响。使用伤残调整生命年等综合指标,可以科学地评价疾病所造成的致死和致残的总负担。整合人口结构与疾病趋势的数据加以建模,可以预测将来疾病谱的改变,尤其在老龄化的大背景下,慢性病的负担会加重,给前瞻性的卫生政策制订给予支撑。成本效益分析可以用来评价不同的干预措施的经济效率,帮助决策者确定防控的重点以及资源的配置方向。

## 2.3 药物不良反应监测

健康大数据技术大大提高了药物安全监测的能力和效率。关联医院用药记录同不良事件报告系统,就能主动察觉药物同不良反应的时序联系信号<sup>[2]</sup>。使用自然语言处理技术挖掘电子病历文本,有利于在临床实践当中发现没有主动上报的疑似不良反应。以大规模医保数据库为基础的真实世界研究,给药物长期安全性评价提供重要的证据。比例失衡分析、贝叶斯网络等智能算法可以对多维数据进行检测,从中发现隐藏的风险信号。根据药物基因组学的信息,可以对不良反应的个体遗传差异进行深入的研究。建立规范的信号评价和因果关系论证程序,是区分真实的药物不良反应和偶然事件的重要方法。

## 3 健康大数据驱动的卫生资源优化配置

### 3.1 医疗服务需求预测

健康大数据给精准预测医疗服务需求赋予了有力的工具。通过历史就诊数据的分析可以得到门诊、急诊、住院量的时间规律和季节性波动。根据人口统计数据以及疾病患病率,可以推算出各个群体、各种疾病所具有

的就医需求。关联气象等环境数据,可以预测某类疾病(呼吸系统疾病)的就诊高峰期。公共假日出行、交通事故等数据也可以用来预测局部医疗压力。利用时间序列模型(自回归移动平均模型等)预测未来趋势,用机器学习算法(随机森林等)将多因素复杂关系加以综合,以提高精度。短期预测可以指导医疗机构排班、物资准备,长期预测可以支撑医疗设施规划、人才培养。

### 3.2 基层医疗服务能力评估

以数据为基础的评价体系有利于客观衡量并提高基层医疗服务水平。用诊疗人次、公共卫生服务完成率等指标统计可以量化机构的工作负荷。用电子病历质控、合理用药率、患者满意度等来对服务质量进行系统的评价。分析双向转诊数据可以反映分级诊疗的实施效果。追踪慢性病管理对象血压、血糖的控制率可以评价长期健康管理的效果。对标分析可以发现机构间的差距以及薄弱环节。创建起包含服务量、质量、效率、满意度的综合评价体系,并且把评价结果同资源分配、激励措施联系起来,就能形成一个持续改进的机制。根据评价结果所发现的能力短板,可以制订出有针对性的医务人员培训和发展计划。

### 3.3 区域卫生规划决策支持

数据分析给科学制订区域卫生规划提供重要支撑。整合居民体检、疾病监测、死因登记等数据,可以对区域健康状况和疾病谱做全面的评价<sup>[3]</sup>。调查统计医疗机构、床位、人员、设备等资源,可以了解供给总量及结构。用地理信息系统分析居民就医的时空可及性,可以找到服务覆盖薄弱的地区。经由比较医疗服务供给同居民健康需求,可判定各个区域的资源供需契合状况。结合人口变化、老龄化、疾病谱变化等趋势做情景模拟,可以前瞻性预测未来资源缺口。使用运筹学的方法对医疗设施的选址和规模进行配置,可以提高资源整体的利用效率以及居民就医的便利性。对不同的规划方案进行成本效益分析,可以给决策提供经济学依据。

## 4 健康大数据应用面临的挑战与对策

### 4.1 数据质量与标准化问题

健康大数据的应用价值高度依赖于其质量与标准化水平。目前,医疗机构信息化建设水平参差不齐,部分基层单位还存在记录不规范、数据采集不完整、关键信息缺失、术语编码混乱等问题。检验结果单位、诊断标准和数据格式的不一致,大大增加了跨机构、跨地区的数据整合与分析难度。数据质量的好坏直接影响分析

模型是否准确,决策是否可靠。因此必须建立系统的数据治理体系,核心就是推行并强制执行统一的元数据标准、数据模型和交换接口规范,从源头上保证数据的一致性和可比性;部署先进的数据清洗、校验和质控工具,对历史和实时数据进行智能化清洗和校正;建立覆盖数据全生命周期的质量监测与评价指标体系,把数据质量表现同有关机构的考核激励挂钩,形成长效的质量改进闭环,为高级分析和深度应用打下坚实的数据基础。

#### 4.2 隐私保护与安全风险

健康数据因为高度的敏感性,它的隐私与安全保护是不能突破的底线。主要困难有数据采集使用过程中知情同意不透明,集中存储有外部网络攻击和内部非授权访问的风险,在实现数据价值挖掘和共享的时候,很难兼顾到个人隐私保护的需求。应对上述问题需要创建起技术、管理、法规三者相融合的综合防护系统。技术上使用同态加密、差分隐私、联邦学习等前沿技术,实现数据可用不可见或者数据不动模型动的安全计算模式。管理方面创建依靠角色和场景开展的细致访问控制、全部操作追溯审计及数据脱敏分级守护制度。法规和伦理上要完善有关立法,确定数据权属、使用边界、各方责任和违规处罚,创建起严格的伦理审查体系,保证所有的数据处理活动合法、合规、合情,在充分保障公民个人信息权益的基础上,推动数据价值的安全释放。

#### 4.3 跨部门协同与数据共享

健康大数据的价值要想得到充分发挥,就必须打破各个部门之间的数据孤岛。数据分散在卫健、医保、民政、公安等系统,其整合遇到管理体制分割、标准不一、权责不明、安全顾虑等障碍。破解之法就是顶层设计和机制创新<sup>[4]</sup>。首先要建立高级别的跨部门协调机制来统筹推进共享工作。从技术角度来讲,要创建起权威统一的政务数据交换平台,强制推行接口、目录、安全等国家标准。管理方面要编制、公布各个部门可以共享的数据资源清单,规定使用条件以及使用权限。可以采用隐私计算、联邦学习等数据可用不可见的技术,在保证安全的情况下实现联合分析。建立合理的利益分配和激励相容机制,通过传染病防控、慢病管理等典型场景来实行试点,得出可以广泛应用的办法。

#### 4.4 专业人才与技术能力

人才短缺是制约健康大数据应用的瓶颈,突出表现为既懂公共卫生业务又精通数据科学的复合型人才严重不足。公共卫生人员缺少数据建模和分析的技能,而

数据科学家对医学知识和业务逻辑又缺乏了解,因此造成了“知识鸿沟”,限制了应用的效果和深度。为此,要构建起多层次、系统化的人才培养体系。在高等教育阶段要推动学科交叉,设立健康信息学等专业。在职人员要进行大量的、长时间的数据技能培训。进一步推进产学研合作,创建实训基地,培养出能够解决实际问题、具有应用性的人才。短期内可以购买服务引进外部技术力量。同时应该开发面向业务人员的低代码、可视化分析工具,降低使用门槛,通过构建知识共享社区、推广最佳实践来提高能力扩散与整体水平。

#### 5 结束语

健康大数据给公共卫生领域带来变革性机会,在疾病监测预警、流行病学研究、资源配置优化等各方面都表现出很大的应用潜力。利用多种数据,采用先进算法,创建智能系统,公共卫生决策由经验主导转变为数据驱动,干预措施由粗放执行变成精准施行。数据质量、隐私安全、部门协同、人才能力等各方面问题要依靠技术创新、制度建设、多方合作共同解决。完善数据治理体系,制定数据质量标准以及质控办法,改善数据可用性。加强安全防护措施,协调好数据利用和隐私保护的关系,保证数据安全合规使用。推进跨部门数据共享,破除信息孤岛,让数据的价值得到最大程度的发挥。加强复合型人才的培养,提高公共卫生人员计算机应用能力。接下来要进一步研究健康大数据在精准公共卫生、智慧健康管理、卫生政策评价等各方面的运用,促进公共卫生服务的智能化和个性化发展,提高人群健康水平,应对重大公共卫生事件。

#### 参考文献

- [1] 陈敏,周彬,肖树发. 健康医疗大数据安全与管理[M]. 人民卫生出版社:2020:709.
- [2] 梁美心. 健康大数据助力医疗卫生领域供给侧改革浅析[J]. 中国乡镇企业会计,2018,(09):296-297.
- [3] 白梨,张家园,魏晓靓,等. 元数据在真实世界医疗环境下数据标准构建中的应用[J]. 中华医学图书情报杂志,2018,27(08):13-22.
- [4] 王冲. 健康大数据在公共卫生领域中的应用与挑战[J]. 健康之路,2017,16(06):272-273.

作者简介:马瑞雪(1987..07-),女,汉族,籍贯:山东省武城县,学历:大学本科,职称:工程师,研究方向:大数据工程技术。