

健康大数据在医疗卫生领域中的应用及挑战

牟忠林, 王雅洁, 陈娟, 张文诗
(海南医学院研究生院, 海南 海口 571199)

【摘要】 健康大数据是医疗卫生领域中非常重要的资源, 合理利用健康大数据的优势服务医疗卫生事业是大势所趋。健康大数据将在医药研发、疾病诊疗、健康危险因素分析、公共卫生应急管理、居民健康管理和精准医疗等方面发挥不可替代的作用。但是健康大数据在医疗卫生领域中的应用也面临着一些挑战, 例如居民隐私数据泄露的风险, 缺少制度机制层面的联动协同, 缺乏更为专业的人才队伍等问题。因此未来应在这些方面加大工作力度, 保证健康大数据在医疗卫生领域的应用中发挥出更大的作用。

【关键词】 大数据; 医疗卫生; 公共卫生; 健康管理

【中图分类号】 R195 **【文献标识码】** A **【文章编号】** 1003-6350(2017)02-0173-04

Application and challenges of medical health big data in the field of health care. MU Zhong-lin, WANG Ya-jie, CHEN Juan, ZHANG Wen-shi. Graduate School, Hainan Medical University, Haikou, 571199, Hainan, CHINA

【Abstract】 Medical health big data is a valuable resource in the field of health care, and making full use of the advantages of health big data to service the medical and health enterprise is the general trend. Medical health big data will play an irreplaceable role in the field of pharmaceutical research and development, disease diagnosis and treatment, health risk factors analysis, public health emergency management, resident health management and precision medicine. However, the application of medical health big data in the field of health care also faces some challenges, such as the risk of data leakage of the privacy of residents, lack of linkage and coordination of institutional mechanisms, lack of a more professional working team. Therefore, much more efforts should be done to protect the privacy of health data and data security, enhance system management, technical innovation and personnel training to ensure that the health big data can play a greater role in the field of medical and health applications in the future.

【Key words】 Big data; Health care; Public health; Health management

2011年5月, EMC(易安信)公司在世界年度大会上首次正式提出了“大数据”(big data)的概念, 随后国内外著名研究和咨询机构开展了一系列相关研究, 从此大数据正式进入公众视野并且开始影响人们的生活。国际数据公司研究报告指出: 2013年全球数据总量超过4.4 ZB, 并且每两年将翻一番, 预计2020年将达到44 ZB^[1]。大数据技术作为看待世界的一种新视角, 能够超越传统的数据统计方法而获得有价值的信息, 更为高效、便捷地服务于经济社会发展, 被称作是引领全球未来繁荣发展的主要技术变革之一。健康大数据又称为医疗健康大数据, 作为大数据系统中的重要组成部分, 是医疗卫生领域中的宝贵资源。科学合理地利用健康大数据, 充分发挥海量的数据规模、多样的数据类型和快速的数据流转等优势是医疗卫生领域发展的大势所趋。

1 健康大数据的主要数据资源

健康大数据, 从理论上来说, 其主要数据资源包括医疗服务的电子健康档案系统数据, 政府和社会在医学研究方面的学术数据, 临床实验数据, 医院医疗保险的费用结算数据, 居民活动、健康管理数据, 政府

公布的人口与公共卫生数据, 医疗供应系统, 公共经济社会生活中网络产生的数据等, 这些数据资源构成了医疗卫生领域健康大数据的初期数据资源^[2]。健康大数据的数据量庞大且类型繁多, 既有大量的结构化数据, 也包含半结构化和非结构化数据。

从实践层面看, 我国的健康大数据资源主要来自医疗卫生领域的信息统计。近年来, 我国医疗卫生行业信息化日趋完善, 建立起了覆盖国家、省、市、县、乡、村的数据统计和上报工作网络, 约有一百万家的医疗卫生机构采用统计直报系统上传相关数据信息, 建立了包括医疗卫生人力和卫生机构、疾病报告与健康监测、卫生资源与卫生服务等不同的动态的大型数据资源库。从20世纪90年代初开始, 我国每5年就在全国范围内开展一次卫生服务调查工作, 并且建立了具有连续性的、覆盖全国各个地区的包含20万人口的家庭和人口基本信息、患病、就医、基本医疗卫生服务等超过200项指标的数据库, 业已形成了可进行时间序列分析的海量数据, 并且数据的增长速度呈现出前所未有的良好态势。县级以上医院基本建立起以患者为中心, 涵盖门诊业务、临床电子病历、检验、影像、

基金项目: 国家自然科学基金(编号: 81260156); 海南省高等教育教学改革研究项目重点课题(编号: HNJC2014-36)

通讯作者: 王雅洁。E-mail: wangyajie323@163.com

医保、远程医疗、教学等医院所有业务流程以及电子图书、电子病案管理、设备物资管理、职能监控等在内的一体化管理信息系统。新型农村合作医疗管理信息系统也已经基本建立,卫生管理部门、新农合经办机构 and 定点医疗机构三者之间已经建立了网络联接和数据资源的共享模式。此外,基本药物监测评价管理系统的建立也实现了对全国各地基本药物与增补药物的品种和采购情况、缺失的基本药物品种数量和销售价格、补偿及报销情况都实现了统计和监管^[3]。

2 健康大数据在医疗卫生领域中的应用

随着我国医疗改革进程不断加快,我国卫生统计工作和信息化建设都得到了显著提升。根据麦肯锡的估算,如果健康医疗大数据得到合理有效的利用,每年可以为美国产生 3 000 多亿美元的财富价值^[4]。我国是世界上人口最多和经济发展最为迅速的国家,如果我们国家充分利用健康大数据,搭建运行高效的数据平台,将为不同的患者、医务工作人员、医疗生产厂商、医疗卫生科研人员、医学教育者和政策制定者提供及时准确的信息,并在医药研发、疾病诊疗、公共卫生政策制定、健康危险因素评估和精准医疗研究等方面创造出更为巨大的经济价值和社会价值。健康大数据在我国医疗卫生领域应用前景将十分广阔,具体表现在以下几个方面。

2.1 健康大数据提高医药研发效率 大数据技术的核心是对搜集到的数据进行专业化的处理,发现隐藏在不同数据之间的相关性。对健康数据加工处理后,也可以对不同患者的疾病、行为或者情绪进行个性化的分析,挖掘患者疾病的特点与其生活习惯之间的关系,并针对患者的疾病特点或症状(靶点)研制出不同的药品。医药研发机构或者公司在药品研发阶段,利用健康大数据技术来分析公众的疾病趋势和药品需求量,并据此确定出更高效率的投入产出比,合理地配置药材资源。除了降低研发成本外,健康大数据还能够帮助医药研发机构或者公司缩短药物的上市时间,增加药物临床试验的成功率,获得市场准入,尽早将更具针对性、更高治疗成功率和更高潜在市场回报的药物推向市场。以往经验发现,使用预测模型后可以帮助医药企业把新药物从研发到推广市场的周期从大约 13 年减少到 8~10 年。使用过程中并能实施有效实时监控。

2.2 健康大数据实现疾病诊疗的科学管理 随着物质生活水平的不断提高,我国居民的疾病谱发生了显著的变化。医院可以通过健康云平台智能采集每位居民的健康数据,并且根据病情为患者安排专家提供专业的咨询服务。专家通过对患者的健康状况做进一步精确诊断后,告知患者未来可能存在的健康隐患,减轻患者的医疗负担,实现疾病预防诊疗的科学化管理。同时医院等卫生机构也可通过远程医疗

系统产生的数据进行实时分析,处理最新数据。利用健康大数据分析医疗资源的使用情况,实现医疗机构的科学管理以及医疗卫生资源的高效配置,提升我国医疗卫生服务水平和效率。此外,大数据还可以提升医疗价值,形成个性化医疗,将生物学数据(如基因、蛋白质组学等相关数据)和患者电子健康病历数据结合起来,使基因测序、个性化药物和患者的疾病诊疗等个性化医疗变成临床实践^[5]。

2.3 健康大数据提高公共卫生应急管理 传染性疾病和重大疫情严重危害着社会公众的生命安全和社会秩序,而健康大数据能够有效的提高公共卫生部门的应急管理。医疗卫生部门可以建立覆盖所辖区域的卫生管理信息平台,收集信息并建立居民的健康信息数据库,利用大数据技术对公共卫生数据进行实时监测和分析,快速检测传染病,对疫情进行全面监测,并通过监测疫情进行预警和处置,这将大大减少医疗支出、降低传染病等疫情的感染率。与此同时,公共卫生部门还可以通过为社会公众提供准确有效的健康咨询服务,提高他们的健康危险意识和卫生防范意识,这也能够在一定程度上降低疫情感染和传播的风险概率。如寨卡病毒,很快在大数据的支撑下分析其结构,得到了有效的防控。

2.4 健康大数据提升危险因素分析水平 通过互联网与医疗卫生信息系统和相关信息系统对社会公众的健康状况进行监测,可以系统全面地获得与健康危险因素有关的数据,再通过大数据技术的深度挖掘分析,找寻出真正威胁公众的健康危险因素。收集的相关数据具体包括:生物因素(包括致病性细菌、真菌、病毒等微生物的监测数据)、环境因素(包括空气、土壤、水文等监测数据)、社会因素(包括收入水平、人口迁移、营养条件、教育就业等监测数据)、心理因素(包括抑郁、自闭、愤怒等监测数据)、家庭遗传因素、卫生服务因素以及个人行为等方面的因素。运用大数据技术对收集到的健康危险因素进行比对和关联分析,针对在不同环境和不同社会条件下生活的人群进行测评,最终确定健康危险因素,并且可以根据确定的危险因素得知特定的疾病容易在哪些特定的生态环境和社会环境中发生。最后针对社会公众的生活领域进行有针对性的干预,促进居民健康水平的提高,同时引导社会公众去有意识地改变和保护环境。

2.5 健康大数据有效地实现居民健康管理 居民电子健康档案是居民健康管理方面的重要基础数据,运用大数据技术进行分析处理,可以为居民提供个体化的健康管理服务,改变传统的营养学和健康学的模式,从环境、营养、社会、心理、运动等不同的方面对不同的居民进行高效的健康服务和管理,有效地帮助和指导社会公众保持身心健康。此外,大数据技术

还可以整合分析患者的健康信息,并且通过分析后的数据为患者进行远程诊断和治疗提供更好的数据证据,减轻患者的心理压力。因此通过大数据技术对居民健康大数据进行分析,能够实现对居民的身心健康的智能化监测,分析影响居民身心健康的因素,进一步帮助居民提高健康管理水平。

2.6 健康大数据为精准医疗提供有效支撑 精准医疗是以个体化治疗为基础,应用基因组、蛋白质组学技术结合患者生存环境、生活方式和临床数据,精确地筛选出疾病潜在的治疗靶点,并根据疾病不同的病理生理学基础将患者分类,最终实现能够针对特定患者制定个体化的疾病预防与治疗方案^[6]。据美国相关研究机构的测算,美国医疗系统每年因不必要和无效果的治疗而浪费 7 500 亿美元,占美国医疗总开支的近三成。而在我国,无效医疗耗费情况则更加严重。如今,精准医疗已纳入我国“十三五”重大科技专项。根据精准医疗计划,我国将重点研发一批国产新型医疗器械和防治药物,并努力形成一批由我国定制并经国际认可的疾病诊疗指南、临床路径和干预措施。精准医疗将为医学发展带来一场全新的变革,改变现阶段的诊断、治疗模式。应该说,精准医疗的研究就是以健康大数据作为有效的支撑,在健康大数据研究的基础上,通过对疾病的精准分类、预防、诊断,为社会公众制个性化、精准化的疾病防治方案,获得疾病发生分子机制的知识体系。其中,随着近年来国家对精准医疗的关注及扶持逐渐上升,癌症液体活检、癌症的个体化疗法等医疗技术身处其中发展迅速。同时,健康大数据还提供具备实时知识产出能力的持续学习型基础设施,开发具有预防性、预测性和可参与性的医疗系统^[7],为药物遗传学、分层医疗保健领域、可穿戴设备和可植入传感器提供数据支撑和技术支持^[8]。大数据分析是精准医学发展的助推器,同时也是跨界合作的重要领域。

3 健康大数据在我国医疗卫生领域应用中的挑战

大数据技术已经深刻改变了人们的生活,并且影响和推动着医疗卫生领域的统计调查和信息化建设工作。科学的运用大数据技术对我国健康数据进行专业化分析处理,能够有效的解决医疗卫生领域曾经难以解决的一些复杂问题。但大数据技术并不是万能解药,如何借助大数据技术更好地发挥健康大数据的作用还存在许多风险及挑战。

3.1 健康大数据可能导致居民隐私数据泄漏 医疗卫生领域大数据的采集工作不可避免地会涉及患者的隐私问题和公众的基本信息,包括患者的病情、部分家庭成员的基本情况甚至基因等遗传数据等。医疗大数据的所有权、管理权、控制权和运营权分别掌握在患者、政府、医院和第三方机构,健康大数据的

应用也需要他们多方的交流和沟通。目前电子病历及其他医院信息系统、医用软件和器材在安全隐私的保护方面依旧缺少具可操作性的有效的标准。目前电子病历及其他医院信息系统、医用软件和器材在安全隐私的保护方面依旧缺少具可操作性的有效的标准。用于科学研究的医疗大数据,尽管都经过匿名化的加工处理,但是仍然存在身份被重新识别和确认的风险,在法规政策上也未形成健全的体制,各合作机构在独自占有数据的前提下,也没有去有效地解决由共享产生的数据安全和数据隐私的问题。近些年发生的一系列私人信息泄露的教训警醒我们,部分旨在保护个人隐私的传统措施,例如告知与许可、匿名化、模糊化等处理手段,在大数据的背景条件下存在失去效用的可能。加之这些医疗数据因保护不力而出现泄露,可能会使患者及家庭成员的生活和工作遭受到不可预知的骚扰^[9]。美国前总统克林顿签署的《健康保险携带和责任法案》就专门为安全与隐私的保护及信息的给你共享制定了策略,为美国电子病例的广泛使用奠定了基础。随着健康大数据广泛应用,信息的可信赖性成为数据共享与交换的前提。我们要在技术上进行强化,更要在政策法规上给予足够的重视。目前加快明文法律法规的制定、清晰对健康数据所有权的定义,同时在管理、物理设施和技术层面加强多系统的保密性、一致性和可用性是突破健康数据交换与共享瓶颈最有效的方法。

3.2 健康大数据缺乏制度机制层面的联动协同 健康大数据是多领域的合作和数据融合,并且数据来源丰富。其中医疗卫生领域中的各项数据,不仅涵盖了医院行政管理和服务结算等方面的数据,同时也涉及到其他海量复杂的数据资源,包括患者在治疗时的临床数据、实验室检查数据和医学影像数据等内容,甚至包括生物学当中的基因学数据以及当地的公共卫生数据。除了以上所提到的数据之外,全社会的其他领域产生的数据资源也可能成为健康大数据的潜在数据来源。我国互联网用户已经超过了 5 亿,包括医疗企业数据在内的万亿 GB 级别的数据每年都在以超过 50% 的速度呈现爆炸式增长,这已经成为医疗卫生领域健康大数据的优势。但是我们同时应该看到,从我国的现实情况来看医疗卫生领域的各类数据目前大多散存在各地区和各部门,如何进行有效的数据融合,如何将有效的融合数据提供给科研协作单位,如何实现跨机构的协作机制,如何促进产学研合作,这就需要医疗卫生相关部门和机构之间通力合作,突破因地区、部门、机构利益造成的壁垒,打破传统的管理机制局限,打通不同业界健康数据的共享渠道,实现数据技术应用的有机融合,为医疗大数据的共享提供一个有利的环境,在制度机制层面实现联动

协同。为了达成这一目标,从宏观上来讲,国家需要在更高层面上进行协调统筹,从法律层面出台相关政策,整合政府、企业、科研机构等各方面数据。从微观上来讲,医院、科研机构和企业都应该提高管理水平,制定出切实而行的规章制度,从内部规范数据的采集和上报工作,有效地实现医疗健康大数据的标准和规范化。

3.3 健康大数据需要实现信息技术层面的创新 医疗卫生领域的数据存在较大的复杂性和特殊性,而普通的文件系统只能提供最底层和最基本的存储能力支持,因此对于健康大数据国家和全社会需要大力开发与之复杂性和特殊性相适应的文件系统。此外,为了改进健康大数据的管理工作,可能还需要在文件系统的基础之上建立数据库系统,然后通过构建索引实现高效的数据查询和筛选等功能,最后通过数据分析技术从健康数据库中提取出对人们有用的知识。从本质上来讲,大数据的支撑技术与基础平台都是属于云计算技术范畴,也就是说云计算技术在卫生领域中对数据资料的存储、管理和分析等功能的支撑是健康大数据应用的前提^[10]。

3.4 健康大数据需要进行更为准确的相关性分析 健康大数据从本质上来讲是将许多不同领域之间的相关数据通过专业化的处理,从中建立起较强的相关性。这就要求医疗卫生统计去收集和利用关系不甚紧密甚至看来毫无联系的数据。从技术层面讲大数据虽然需要获取全部的信息数据,但是在一些情况下可以允许少量数据不精确或者丢失。因为大数据技术与之前相比已经大幅度的提高了统计分析的精准度,但是健康大数据的半结构化和非结构化的数据在统计分析时具有较高的难度,需要进行更为准确的相关性分析。因此,为了提高健康大数据的可靠性,这就对包括医院在内的医疗卫生服务机构提出了更高的要求,对于数据处理和上报时要注意实时性和准确性,医疗卫生领域的统计调查研究部门和医疗卫生机构应该引起足够的重视。

3.5 健康大数据的应用缺乏专业的人才队伍 健康大数据和大数据技术的有效应用是建立在医疗卫生服务系统高度信息化的基础之上,而现实情况是国内医疗卫生信息化水平较低,无法完全满足和适应大数据技术的要求。国内医疗卫生信息化水平较低主要有两个原因,一是因为我国还有部分地区医疗卫生硬件设施较为落后,无法开展健康数据的采集和上报工作;另外一个更重要的原因是因为卫生部门和医疗机构严重缺乏专业技术人员。由于我国的健康大数据的应用还处于起步阶段,整个医疗卫生行业非常缺乏既熟悉医疗业务又擅长信息技术的高水平复合型人才,加之许多医疗机构的人才流动性较大,很难形成有效的知识积累,进而影响信息化进程。当代社会的竞争是技术和人才的竞

争,归根结底是专业人才的竞争,如果医疗卫生行业长期缺乏既熟悉医疗业务又擅长信息技术的人才,那么我国医疗卫生领域的大数据技术就很难发挥真正的作用。因此国家教育部门需要逐步提升医学信息技术人才的能力。除此之外加快开展医学专业人才培养工作也是时代所需,只有将医学专业人员和信息技术人才结合起来,才能更好地适应医学大数据时代的发展,更好地面临大数据时代的挑战。

4 结 论

医疗卫生事业作为改善我国全民健康的重要基础和保障,在近些年的高速发展的同时累积了我国民众的各式各样的数据资源,这些数据资料连同全社会其他相关的数据信息构成了我国的健康大数据,健康大数据的建立将会推动我国医疗卫生事业的发展。健康大数据在我国医疗卫生领域中的应用前景十分广阔,具体表现在医药研发、疾病诊疗、健康危险因素分析、公共卫生应急管理、居民健康管理和精准医疗等方面。但是发挥健康大数据的作用还面临着许多挑战:一是健康大数据可能面临居民隐私数据泄露的风险,二是需要实现制度机制层面的联动协同,三是需要实现信息技术层面的创新,四是进行更为准确的相关性分析,五是需要培养更为专业的人才队伍。因此下一步应在健康大数据隐私保护和数据安全、制度管理、技术创新和人才队伍培养等方面加大工作力度,保证健康大数据在医疗卫生领域的应用中发挥出更大的作用。

参 考 文 献

- [1] Terner V, Gantz JF, Reinsel D, et al. The digital universe of opportunities: rich data and the increasing value of the internet of things [R/OL]. IDC Analyze Future, 2014. [2014-05-10]http://idcdocserv.com/1678.
- [2] 郑玲微. 大数据来临,你准备好了吗——步跨大数据时代[J]. 信息化, 2013, 1: 10-13.
- [3] 周光华, 辛英, 张雅洁, 等. 医疗卫生领域大数据应用探讨[J]. 中国卫生信息管理杂志, 2013, (4): 296-300.
- [4] Manyika J, Chui M, Brown B, et al. Big data: the next frontier for innovation, competition, and productivity [R]. McKinsey Global Institute, McKinsey&Company, 2011.
- [5] 许德泉, 杨慧清. 大数据在医疗个性化服务中的应用[J]. 中国卫生信息管理杂志, 2013, 10(4): 301-304.
- [6] 巩鹏. 对中美精准医学差异及我国精准医学发展的思考[J]. 医学与哲学, 2016, 37(8): 26-27.
- [7] Hood L, Flores M. A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory [J]. New Biotechnol, 2012, 29(6): 613-624.
- [8] Yang GZ. Body Sensor Networks. 2nd ed. London: Springer-Verlag, 2014: 5-7.
- [9] 沈韬, 崔泳. 医疗大数据: 期望与现实[J]. 中国数字医学, 2015, 10(7): 2-4.
- [10] 颜延, 秦兴彬. 医疗健康大数据研究综述[J]. 科研信息化技术与应用, 2014, 5(6): 3-16.

(收稿日期:2016-12-07)