

国家重点临床研究中心下的健康医疗 大数据平台研究建设与发展

屠强¹, 徐冬², 平措³

1. 中国科学技术大学附属第一医院西区 安徽省肿瘤医院信息科, 安徽 合肥 230001;

2. 中国科学技术大学附属第一医院 安徽省立医院信息中心, 安徽 合肥 230001;

3. 山南市人民医院信息科, 西藏 山南 856000

摘要: 随着医疗健康产业的快速发展, 国家临床研究中心下的健康医疗大数据平台建设迫在眉睫。本文基于临床研究中心研究背景, 通过对国家级临床研究中心下的健康医疗大数据平台的构成部分进行说明分析, 对比现有的医疗数据平台分析其可行性以及临床研究中心下的健康医疗大数据平台的关键技术, 并结合当下大数据快速发展的趋势, 总结建设国家级临床研究中心下的健康医疗大数据平台的建议和对策。

关键词: 大数据; 医疗健康; 临床研究中心

中图分类号: R 197.32 **文献标识码:** B **文章编号:** 1674-8182(2019)05-0700-04

Construction and development of the big data platform for health care under the National Clinical Research Center

TU Qiang*, XU Dong, PING Cuo

*Department of Information, The First Affiliated Hospital of USTC West District

(Anhui Province Cancer Hospital), Hefei, Anhui 230001, China

Abstract: With the rapid development of the medical health industry, it's urgent to build a large data platform for health care under the National Clinical Research Center. Based on the research background of the Clinical Research Center, this paper illustrates and analyses the components of the big data platform of health care, compares the feasibility of the existing medical data platform and the key technologies of the big data platform of health care under the Clinical Research Center. And the trend of rapid development of big data at present was combined to summarize the suggestions and countermeasures of the big data platform of health care under the National Clinical Research Center.

Key words: Big data; Medical health; Clinical Research Center

Fund program: Medical Research programs of Anhui Health and Family Planning Commission (2005006); Anhui Science and Technology program (11010202193); Strategic Emerging Industries programs in Anhui Province (11010401016)

随着信息技术的突飞猛进,“得网络者得天下”的时代已经来临。国家“互联网+”战略推动着各行各业掀起数字化浪潮,大数据在医疗领域的应用必然会极速普及。搭建大数据平台不仅有利于提高医院管理精细化水平,降低医生疾病诊断误判率,提升医院服务效率,还在增强医院综合竞争力等方面起到了关键作用。值得注意的是,大数据在应用的同时必须有效预防患者信息被泄露和隐私权受到侵害^[1]。

国外对于医疗大数据的研究起步较早,研究的理论深度都较为成熟。通过梳理文献,发现国外学者大

多侧重于大数据技术研究方向,其关注的重点在于如何把大数据应用到提高医疗服务中^[2],而对于临床中心积累数据的分析却鲜有研究。因此,本文侧重研究医院临床中心依据医疗数据的深入分析,为建设国家级健康医疗大数据医平台,提供准确的数据依据。

近年来,各医院中以临床研究中心应用为导向的信息化建设步伐加快,大数据分析的思想、方法与技术也得到了蓬勃发展^[3-5]。采用大数据技术助推临床研究的条件日渐成熟,客观上为研究提供了基础。在我国医学领域前沿技术需要重点突破的领域中,临

床研究中心下的健康医疗大数据分析是重中之重^[6-8]。在我国医学界,对医疗领域大数据使用的研究,主要集中于政府推动的大数据医疗保健方面。这与促进大数据医疗的实际情况相符合。但对临床研究中心实验数据、疾病诊断数据、居民行为健康数据等汇集形成的医疗大数据呈现的数据规律研究却欠佳^[9-10]。这一状况,反映出医疗卫生领域“等、靠、要”的问题,由于大数据在医疗领域的应用长期形成的习惯,政府顶级设计并选择特定区域试点。待发展成熟后,再向全国范围内推广。

1 临床研究中心下的健康医疗大数据平台组成特点

由于医疗行业的特殊性,医疗卫生行业的数据呈现出以下特点,同时由于医疗信息的储存、采集等历史方面原因,也会导致其异质性和关闭性问题。

1.1 数据量大 医院一般有多个临床研究中心信息系统,建设跨临床信息系统的统一临床数据中心,将各临床研究中心数据汇聚到一个统一数据库中,并在统一平台上关联性可视化显现,以实现这些临床数据的集约化应用,数据量成倍增加。截至 2011 年,美国卫生系统的数据量达到 150 EB。除了传统临床和测试产生的数据外,医疗和卫生机构还开发连接所有互联网对象,例如便携式医疗设备上的二维代码标签生成的数据,随着 Internet Things 技术的发展,数据量要大得多。信息收集和分析的速度由原来的“天”到“小时”,再到现在的“秒”来计算。此外,遗传数据也是一个巨大的实体,全面的基因测序,并为个人用户产生各种个人健康数据。

1.2 复杂度高 医疗领域包含许多医学术语,不仅包括 30 000 种疾病的数据,还包括非结构化数据,如成千上万的诊断,手术和药物名称,大量图像和医疗指示。由于医疗数据是不同临床研究中心服务的产品,基于临床研究中心的数据统计研究目前公认的是采用大样本随机对照试验方法。医院之间存在许多差异(如患者的人格和疾病水平、医院诊断和治疗水平、医疗数据记录和编码水平)。即使使用相同的描述格式,其语法和语义也不相同,这会导致数据复杂性。

1.3 完整性强 把大数据应用于医疗卫生领域,已成为共识。目前,以单家医院的临床研究中心数据规模而言,尚不能称之为大数据。因此,数据完整性和约束完整性在数据处理过程中保证了限制。数据完整性是指数据的准确性,统一性和可兼容。

1.4 安全性好 除了有关患者隐私的医疗数据外,医疗数据还包含大量有关医院运作,诊断和治疗方法

以及药物疗效的信息。这些信息通常很敏感,有些可能涉及商业利益。因此,大多数医疗机构不愿意披露数据,导致一些可以处理数据的部门没有数据。在建成医院统一临床数据研究中心基础上,建立各临床专科、亚专科数据仓库,就可有效地保护原始数据不被篡改和患者隐私的医疗数据不被泄露。

1.5 多样性显 异质性问题 导致了各种各样的数据,主要包括数据来源的异质性,管理系统的异质性和所用标准的异质性。综合医疗服务平台拥有广泛的数据(包括医院,外部测试组织,新医疗合作医疗保险,医疗保险,个人用户和网络等),费用补贴/使用,健康管理/社交网络;平台数据内容多种多样(包括病史、身体检查、理化检查、居民基本健康记录、各种个人信息和网站等)。涉及的数据源类型多种多样,包括结构化数据、半结构化和非结构化数据;异构管理系统具有在不同管理系统,数据库和使用不同技术的不同管理系统上运行的不同操作系统。为实现异质性,目前使用的标准包括 CDA、HL7、DCOM 等。

上述五大特点,是基于打破信息孤岛而实现信息数据标准化基础上呈现的。因此,要把传统信息孤岛打破,就必须制定统一的数据标准或表示方法,并统一封装,以确保统一的处理。以使在全国各独立医疗系统之间能够互联互通。

2 国家级临床研究中心下的医疗健康大数据平台建设的可行性及关键技术分析

未来将与公民、医疗保险机构、医疗检查站、第三方检查/成像、积极的医疗服务和新的农村医疗合作,在整合社会医疗资源基础上,建立综合性的区域健康医疗生态系统平台。制定健康医疗数据标准,实现统一标准的大数据集成为协作。技术支持层面上用分布式计算平台进行大数据处理。目前,医疗健康数据平台通常建立在 Hadoop 框架之上。理论上,平台技术系统可以处理集成医疗生态系统中的各种类型的数据,实现一平台信息数据存储和大规模信息处理。Hadoop 可以提供具有系统低级细节的分布式透明基础架构,以及以可靠、高效和可扩展的方式处理大量数据的分布式软件架构。

Hadoop 应用程序与现有行业之间仍存在差距。对于医疗应用,主要有两个方面:首先,没有分析库和算法库以及医学领域的模型库;第二,没有医学知识库。在医学健康数据应用等临床决策支持下,需要了解医生输入数据的内容,做出必要的决策支持。数据内容的语义处理必须基于基础医学知识。

基于医院临床研究中心的大型医学数据平台要

解决的关键技术:医学健康语义知识的基本构建,语义数据语义处理,传统分析与挖掘技术的协调。

基于医院临床研究中心的大型医疗数据平台的构建应考虑如何利用现有数据和收集哪些数据。还要考虑 Hadoop 中原始联系人数据与非联系人数据之间的关系,以及兼容的业务数据和外部数据。

在技术层面的分析研究开发上面,研究医学大数据应用的需求,其次,研究医学大数据集成和融合技术,数据语义处理技术,大数据分析和挖掘并行处理技术等关键技术;核心集成互联网数据,政府相关行业数据,第三方健康数据(PHR),个人遗传数据和生物样本数据,完成大型医疗数据中心的建设,建立基于此的医疗大数据平台,开展医学研究分析,实施基于大数据的医疗保健应用。

3 面临的困难及挑战

3.1 数据上传的范围、质量把控以及分析模型的建立 大数据是一个新兴的信息化术语。许多医疗机构管理人员和医务人员都不了解大数据的应用场景和价值,因此很难提出对大数据的确切需求。为此,医疗保健系统管理部门与研究所合作。汇集大数据从业者和专家,推动和共享大数据应用场景,让更多的业务人员了解大数据的价值,并提出数据分析的需要,从而推动数据挖掘的方向和发展。同时,医院管理方,医院实施方和数据上传方应严格遵守数据管理准则,数据高质量保证来自精准的数据分析。

3.2 云平台的安全问题 灵活性、共享性的云计算优势为医疗资源的区域发展提供了新的机遇。然而,共享和经济健康云平台,存储的数据涉及患者的隐私,如果泄露,后果不可想象。但是,在构建过程中,信息安全问题会更加复杂,因此需要对云平台和各种应用系统进行全面的安全保护。

3.3 专业技术人才缺乏 医疗大数据平台的建设是一个复杂的系统工程需要大量的多方面专业人才来共同解决,目前大多数医院缺乏医学信息人才,缺乏高素质的综合信息技术系统开发人才,严重制约健康医疗大数据平台的建设。因此,需要出台相关政策,吸引相关人才汇聚,共同开发。

3.4 数据是否保持高质量 目前,医疗数据来源于不同的医疗机构,包括各级医疗机构,社会福利中心,疾病预防控制中心和预防性健康管理机构等。各类医院采用不同开发商所提供的大数据收集系统。这些系统旨在满足部门或部门的业务需求。不正确的数据和标准不一致的数据等数据乱象,严重降低了数据使用价值。在临床研究中心大数据建设中,必须建立以

患者主索引为线索的数据组织架构,实现对患者历次临床事件(门、急诊,住院等)涉及的所有临床数据的跟踪。

3.5 数据标准化与否 由于不同医疗机构的信息水平参差不齐,加之一些医疗机构的数据信息系统来自不同开发商。各开发商用的数据系统标准又不同,这实际多达 200 多个标准,就使得各系统的兼容性无法保证,无法通过系统构建轻松解决问题。只有基于统一平台的数据集成的系统,才可以真正实现标准的登陆和数据互联。

4 对策与建议

4.1 加强临床研究中心医疗大数据平台建设 医疗服务需求和新时代的推拉,倒逼医院必须加强大型医疗数据平台的建设,就要解决医疗大数据平台建设中客观上存在诸多的技术问题。医院要积极提出健康医疗大数据平台建设的时代事业需求,积极推进政府决策和政策支持,并最大限度地提高第三方机构参与大型医疗数据建设的力度。

4.2 注重临床研究中心医疗大数据安全保障 在建立基于医院临床研究中心的大型医疗数据库的过程中,医院应注重隐私保护,知识产权,数据披露和信息安全。要在系统平台开发上,把安全端口前移。通过建立安全系统,建立全面的科学数据保护体系。

4.3 加强临床研究中心医疗信息化复合型人才队伍建设 临床研究中心医疗信息化复合型人才队伍建设,难以一蹴而就。应该多渠道引领,多院校培养,实践中锻炼。

5 总结

在大数据时代,基于临床研究中心下的大型医疗数据信息化建设已成为必然趋势,这是医院患者发展的必然要求,所有医院工作的顺利开展需要建立在大数据平台的基础之上。通过云计算处理模式,可以及时获取相关信息,从而可以提高科学决策水平。同时,基于临床研究中心下的大型医疗数据的构建应从数据访问控制入手,为集成数据管理信息创造良好环境,从而提高综合数据管理水平。因此,国家重点临床研究中心下的健康医疗大数据平台建设与发展迫在眉睫。

参考文献

- [1] 蒋言斌,麻欣张乐,李响.论医疗大数据患者隐私权的请求权[J].医学与哲学(A),2018,39(6):50-52,60.

- tients after lung cancer surgery[J]. *Interact Cardiovasc Surg*, 2013, 16(6):819-823.
- [3] 林琳, 戢艳丽, 车国卫, 等. 肺癌胸腔镜肺叶切除术后不同药物镇痛效果的随机对照试验[J]. *中国胸心血管外科临床杂志*, 2017, 24(11):830-834.
- [4] 潘月莲. 肺癌胸腔镜肺叶切除术后患者疼痛的评估及护理[J]. *实用临床护理学电子杂志*, 2018, 3(4):139-140.
- [5] 黄明明. 规范化疼痛管理在肺癌患者围手术期的研究[D]. 昆明: 昆明医科大学, 2017.
- [6] 张高峰, 孙立新, 陈怀龙, 等. 不同镇痛方式在胸腔镜肺叶切除术患者术后镇痛效果的比较[J]. *临床麻醉学杂志*, 2014, 30(10):984-988.
- [7] Barr J, Fraser GL, Puntillo K, et al. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit[J]. *Crit Care Med*, 2013, 41(1):263-306.
- [8] Jablonski J, Gray J, Miano T, et al. Pain, agitation, and delirium guidelines: interprofessional perspectives to translate the evidence[J]. *Dimens Crit Care Nurs*, 2017, 36(3):164-173.
- [9] 沈曲, 李峥. 休斯顿疼痛情况调查表信度和效度的研究[J]. *中华护理杂志*, 2006, 41(11):1049-1051.
- [10] 车国卫, 刘伦旭, 石应康. 加速康复外科临床应用现状与思考[J]. *中国胸心血管外科临床杂志*, 2016, 23(3):211-215.
- [11] Schug SA, Palmer GM, Scott DA, et al. Acute pain management: scientific evidence, fourth edition, 2015 [J]. *Med J Aust*, 2016, 204(8):315-317.
- [12] 王峻, 王春莉, 侯丽莉, 等. 创伤骨科无痛病房疼痛管理模式的建立与评价[J]. *护理研究*, 2014, 28(21):2668-2670.
- [13] Foss M, Bernard H. Enhanced recovery after surgery: implications for nurses[J]. *Br J Nurs*, 2012, 21(4):221-223.
- [14] 湛永毅, 卿利敏, 刘翔宇. 基于 JCI 标准的肿瘤患者疼痛管理实践[J]. *中国护理管理*, 2015, 15(4):491-494.
- [15] 温彩霞, 徐玲芬, 吴巍. ICU 成人机械通气患者实施镇静镇痛策略的研究进展[J]. *临床护理杂志*, 2018, 17(1):59-62.
- [16] 寇任重, 范刚启, 刘岚青. 基于多学科协作的头痛单元组建与运作分析[J]. *中国医院管理*, 2017, 37(6):36-38.
- [17] 车国卫, 李为民, 刘伦旭. 快速肺康复需要围手术期流程优化[J]. *中国胸心血管外科临床杂志*, 2016, 23(3):216-220.
- [18] 张银文, 潘亚男, 毛晓博. 快速康复外科指导下胸腔镜手术围术期处理的重建[J]. *中国医药导报*, 2017, 14(8):115-119.
- [19] 漆奋强, 袁天柱, 吴根社, 等. 电视胸腔镜肺叶切除术在老年肺癌患者中的应用及对代谢反应、疼痛应激与心肺运动耐力的影响[J]. *中国老年学杂志*, 2018, 38(10):2353-2355.
- [20] 车国卫. 加速康复外科——人文 or 技术? [J]. *中国肺癌杂志*, 2018, 21(03):168-172.
- [21] 时辉, 梅龙勇, 车国卫. 肺癌术后胸腔闭式引流术应用的新观念[J]. *中国肺癌杂志*, 2010, 13(11):999-1003.

收稿日期:2018-10-25 修回日期:2018-11-10 编辑:王宇

(上接第 702 页)

- [2] 田海平. 生命医学伦理学如何应对大数据健康革命[J]. *河北学刊*, 2018, 38(4):33-41.
- [3] 王艺, 任淑霞. 医疗大数据可视化研究综述[J]. *计算机科学与探索*, 2017, 11(5):681-699.
- [4] 薛付忠. 健康医疗大数据驱动的健康管理学理论方法体系[J]. *山东大学学报(医学版)*, 2017, 55(6):1-29.
- [5] 王灵芝, 郝明. 医疗大数据的特征及应用中的伦理思考[J]. *医学与哲学(A)*, 2017, 38(4):32-35.
- [6] 崔娜. “医疗大数据应用技术国家工程实验室”正式启动[J]. *北京大学学报(医学版)*, 2017, 49(6):1102.
- [7] 卞伟玮, 王永超, 崔立真, 等. 基于网络爬虫技术的健康医疗大数据采集整理系统[J]. *山东大学学报(医学版)*, 2017, 55(6):47-55.
- [8] 马灿. 国内外医疗大数据资源共享比较研究[J]. *情报资料工作*, 2016(3):63-67.
- [9] 沈沂. 医疗大数据对医院图书馆文献信息服务的影响[J]. *图书情报工作*, 2016, 60(S1):33-36.
- [10] 何秀丽, 任智源, 史晨华, 等. 面向医疗大数据的云雾网络及其分布式计算方案[J]. *西安交通大学学报*, 2016, 50(10):71-77.

收稿日期:2018-11-20 修回日期:2018-12-25 编辑:王国品