

医院开展医学大数据应用初探^{*}

吴云霞 刘忠军 张振会

李 维 张 晨

(北京大学第三医院骨科 北京 100191)

(北京大学第三医院信息管理中心 北京 100191)

〔摘要〕 阐述医院医学大数据的来源、特点以及在医学研究实践中的主要应用,指出医院医学大数据应用场景广泛,但同时面临众多问题与挑战,以期在医院医学大数据应用提供借鉴。

〔关键词〕 大数据;医学科研实践;应用研究;随访数据;文本挖掘

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.12.004

Preliminary Exploration of the Application of Medical Big Data Implemented by Hospitals WU Yunxia, LIU Zhongjun, ZHANG Zhenhui, Department of Orthopedics, Peking University Third Hospital, Beijing 100191, China; LI Wei, ZHANG Chen, Information Management Center, Peking University Third Hospital, Beijing 100191, China

〔Abstract〕 The paper dilates upon sources, features of medical big data of hospitals and its main application in medical studies and practices, points out that medical big data of hospitals having extensive application scenarios, is confronted with multiple problems and challenges, in the hope that it would provide the application of medical big data of hospitals with references.

〔Keywords〕 big data; medical scientific research practice; application study; follow-up data; text mining

1 引言

随着移动互联网、人工智能、物联网的普及,全球数据总量呈现指数级增长态势。与此同时数据科学日趋活跃,促进大数据的理论与技术在各领域广泛应用。目前我国大数据发展在政策、学术、研发、产业等多个层面都取得显著进展。以医学大数据为例,互联网医疗及智慧医疗在政策有力推动和技术日趋成熟下不断取得新成就,促进医学领域不

同学科、专业的融合,发挥交叉学科的新功能、新作用。本文结合大数据^[1]及医学数据的特点,总结医学大数据的内涵:(1)大量(Volume)。包括全体住院患者数据,即患者的全息数据(患者从第1次挂号至痊愈全过程)。(2)高速(Velocity)。通过电话、微信、短信、邮件等各种方式与患者进行互动,实现快速、实时收集患者数据。(3)多样(Variety)。医疗数据包含多种类型、格式(如数字、文本、影像、音频及视频等格式)、来源(如电话、短信、邮件及微信等)、时间(如从挂号至痊愈期间在诊断、治疗、康复及随访各时间点的数据)、空间(如利用现代化的通讯方式通过各种方式联系到异地患者实时采集信息)、渠道(从门诊、急诊、转诊等各种方式就诊的患者)。(4)价值(Value)。分析并挖掘医疗过程中的数据、发现问题、总结及预测发病、治病规律,为提高患者满意

〔修回日期〕 2018-10-12

〔作者简介〕 吴云霞,副主任技师,发表论文3篇;通讯作者:刘忠军,教授,主任医师,博士生导师。

〔基金项目〕 原卫生部办公厅2010年度国家临床重点专科建设项目。

度、医院诊疗能力和制定行业政策提供数据支持。本文通过具体案例,重点阐述大数据在医院医学研究中的应用、挑战及问题,以期在医院医学大数据应用提供借鉴。

2 医院医学大数据来源

本研究基础数据来自北京大学第三医院骨科的纸质病历和电子病历,资料收集范围包括门急诊、住院及随访的全息数据收集、整理及分析。收集方式分为 3 种:一是纸质病历,需要逐页逐项录入。二是 2016 年骨科完成随访数据实时收集,患者可通过在固定门诊随访或用手机、短信及微信等方式与医院实时互动,目的是方便患者填写各种生活质量评分,上传人体像,到医院来只是完成必要的医疗诊疗,这样使患者大大节约就诊时间,提高随访率,减少医院无效人群,使医疗资源得到更合理的应用。三是运用文本挖掘技术从大量、无结构的文本信息中发现潜在的数据模式、内在联系、规律、发展趋势等,抽取有效、新颖、有用、可理解、散布在文本文件中的有价值知识,利用这些知识更好地组织信息^[2]。骨科标准化电子病历已于 2018 年 8 月运用文本挖掘技术完成数据结构化,截至 2018 年 8 月 27 日共结构化数据 1 588 138 例,其中门急诊数据 1 523 573 例,住院数据 64 565 例,文本挖掘数据从 2010 年至当前实时病历,其中自 2012 年 1 月 1 日后为全部实时显示。

3 应用示例

3.1 分析发病率的人群特征与地域特征

脊椎退行性疾病(以下简称脊退)主要包括颈椎、胸椎、腰椎的间盘突出及椎管狭窄等病(以下颈、胸、腰椎退行性疾病简称颈退、胸退及腰退)。国内外文献中经过 ICD10 编码、三级骨科医师确认,接受过手术的患者的人群特征分析比较少。“2008-2014 年北医三院骨科脊柱退行性疾病的住院人群特征分析”一文中,确诊患病数据共收集 22 951 例,颈退 10 941 例(47.7%),胸退 588 例

(2.6%),腰退 11 422 例(49.8%)。颈退中男性占比 63.9%,女性 36.1%,高发年龄在 50~54 岁之间,办公室人员占比 24.85%,发病占比最高^[3]。分析脊退伴随疾病,其中高血压占 20.7%、糖尿病占 9.7%,这两种疾病曾被认为在脊退中发病率高,但通过大数据聚类分析发现其发病率低于正常流行病学调查数据,高血压占 24%^[4]、糖尿病占 11.6%^[5]。颈退伴随疾病位韧带肥厚或骨化占 86.1%。在腰退伴随疾病膝关节炎占 70.0%、骨质疏松占 86.1%,其中膝关节炎 85.0%、骨质疏松 91.1%集中在女性。上述这些具体数据为下一步制定精准的研究方案提供依据。但同时该篇文章中的数据 76.4% 来自中国华北地区,95.3% 为汉族,表示此数据存在地域和种族的局限性。

3.2 从卫生经济学角度分析自主创新技术价值

2016 年发表的“对采用新技术治疗寰枢椎疾病治疗费用变化的分析”一文从卫生经济学角度分析自主创新技术治疗寰枢椎疾病 10 年间的费用变化^[6]。2005-2014 年运用自主创新技术共治疗寰枢椎患者 1 489 例,患者例数不多,但时间跨度为 10 年。从经济学角度分析疗效,2005 年每例患者平均收费是 20 851 元,2014 年为 20 878 元;住院时间由 2005 年的 20.57 天下降至 2014 年的 10 天,住院时间显著下降;同时其出血少,疗效好,运用我国自主创新技术治疗费用远低于国际产品。根据国家统计局数据,我国居民消费指数 2005-2014 年累计增长 31%^[7-8],实际住院费用下降 23.3%。根据 2014 年《中国卫生统计年鉴》数据,以 2005 年为基础,我国医药费开支每年上涨 17.49%,至 2014 年已累计上涨 320%^[9]。此外有数据显示手术患者的人工手术费用 10 年未增加,表明医务人员人工费用过低,值得相关部门关注^[6]。

3.3 分析住院时间变化

颈退性别男女比由 1994 年 2.81 至 2014 年 1.61,男性患者依然占比高,而女性患病率相对增加;住院时间由 1994 年 42.35 天至 2014 年 6.1 天,技术更熟练、设备更先进、管理效率增强等因素使

住院时间极大下降; 年龄由 1994 年 50.4 岁至 2016 年 52.4 岁, 但随着我国国民老龄化速度的增快, 颈椎病发病呈年轻化还是老龄化需要进一步深入研究^[10]。

4 思考

4.1 应用场景广泛

4.1.1 概述 医院医学大数据潜在利用价值极高, 在医院管理、预防医学与个性化精准诊治等研究中提供参考与支持, 进而提高医疗机构、医院、医生的效率, 提升诊治准确度。

4.1.2 掌握医学全息数据 从医疗、教学、科研、运营及管理等方面进行全面的数据梳理, 全局数据是下一步进行数据分析与挖掘的重要基础。

4.1.3 发挥决策作用 研究大数据在某一医学领域的应用, 首先要做好全面的顶层研究设计, 建立长、短期研究规划, 为顺利地开展下一步工作提供可行性目标决策。

4.1.4 辅助临床决策、疾病诊断 大数据的出现为临床诊断、决策带来新的发展机遇, 不仅能够通过智能诊断减少错误发生率, 还能够不断发现病情之间的隐含联系, 开创疾病诊断研究的新领域。

4.1.5 远程患者实时监控 为预防疾病或突发事件的发生, 可对高危人群以及随访人群等进行实时监控, 做到防患于未然。

4.1.6 慢病及健康管理 一些大数据平台在对慢病患者、慢病高危人群、健康个体进行健康医疗数据收集和监测的基础上通过深度学习、数据挖掘等技术建立慢病预测分析模型, 找到慢病发生的高危因素, 从而对健康个体给予相关指导, 以及对慢病患者进行个性化治疗与愈后跟踪监测。

4.1.7 预测作用 2009 年《自然》杂志发表一则成功的流感预测案例^[11], 如今利用海量的医疗数据进行疾病风险预测, 和传统方法相比能够提高多种重大疾病、流行疾病预测的准确性。

4.1.8 精准医疗与个性化治疗 通过收集患者的全息数据, 结合其生活方式和环境, 可实现更加精准的诊断, 从而制定个性化的诊疗方案。

4.1.9 科研指导作用 大数据推动各类临床科研的发展, 使临床科研针对性更强。

4.2 问题与挑战众多

4.2.1 问题 尽管医院医学大数据应用场景广泛, 潜在价值巨大, 但必须认识到目前医院医学大数据应用过程中存在的局限与问题: (1) 数据不是万能的, 科学研究以临床定性经验为前提更有效。(2) 不要一味追求“数量大”, 数据的“大”不仅是指数据体量大, 时间跨度大、空间维度广、研究领域跨度大都可以代表大数据中的“大”。(3) 不要盲目追求高级数据模型, 在没有定性经验的前提下运用各类高级数据分析模型并嵌套使用的临床价值产出概率低且周期长。(4) 医学大数据应用过程中面临着数据内容不全、精细化不够、逻辑性不强等问题, 是制约良性健康发展的一大重要因素。(5) 数据之和的价值远远大于数据的价值之和, 而我国医疗数据普遍不能互通互认, 直接导致各医疗机构大量有价值的数据变成数据孤岛。

4.2.2 需关注的方向 应认识到医学大数据研究与应用中需要关注的几点: (1) 统计学是大数据的核心技术支撑, 建立大数据统计分析的新理论和新方法, 发展适应于大数据的统计算法, 能够为大数据技术发展和大数据行业应用提供有力支撑。(2) 人才资源是保障, 医疗大数据研究需要专业人才的支撑, 尤其是具备生物医学、统计、数据分析能力的复合型人才, 而目前这种能够将技术、业务、数据分析结合起来的复合型人才缺口较大。(3) 医学大数据的发展与应用离不开政府、社会及医院等各界的支持, 且医院提供良好、有效的数据整合平台是实现医疗大数据的重要前提, 更有助于实现数据价值的最大化。

5 结语

随着我国社会整体信息化程度不断加深, 以云计算、物联网、移动互联网和人工智能为代表的新一代信息技术对健康医学事业的革命性影响日趋明

(下转第 32 页)

为 7~35 天左右,而现在只需要 24~48 小时甚至更短时间就能实现商业保险理赔报销,减少患者在医院或保险公司排队的时间;避免患者出院排队缴费、后续理赔金申请的繁琐,解决以往商保患者理赔难、理赔慢的问题,为患者提供更加便利高效的服务。对于医院而言,为就医患者构建便民服务体系,有效提升医院服务质量,提高医疗机构工作和管理效率,是顺应国家卫生事业主动服务患者发展趋势的重大举措,为推动“三医联动”、医保改革发挥重要作用。

医院信息化建设依托于医院业务发展需求,随着商业保险和医疗服务不断深入合作,医院信息基础和数据平台建设越来越重要,通过信息化建设实现医院与各类医疗产品和服务融合,全面提升医院综合服务水平,提高居民就医满意度。商保线上快速理赔平台的搭建将医院优质的医疗资源与商保客户的健康需求紧密结合起来,不断推动和促进传统医疗模式和健康管理模式的深刻变革,为越来越多的商保患者实现便捷就医创造更好条件,助力健康中国。

(上接第 20 页)

显。应将医学大数据予以充分利用,从而更好地服务于人们健康生活。

参考文献

- 1 维克托. 迈尔-舍恩伯格, 肯尼思. 库克耶. 大数据时代 [EB/OL]. [2018-08-31]. <https://baike.so.com/doc/5340323-5575766.html>.
- 2 Yanli Xu, Rong Zhao. The Literature Review of Text Data Mining [J]. Science Discovery, 2017, 5 (6): 438-443, 442.
- 3 吴云霞, 刘忠军, 刘晓光, 等. 2008-2014 年北医三院骨科脊柱退行性疾病的住院人群特征分析 [J]. 中国脊柱脊髓杂志, 2016, 26 (1): 70-76.
- 4 高血压联盟(中国), 国家心血管病中心, 中华医学会心血管病学分会, 等. 2014 年中国高血压患者教育指南(简明版) [J]. 中国循环杂志, 2014 (29): 131-140.
- 5 陆菊明. 中国 2 型糖尿病防治指南(2013 年版)更新要点的解读 [J]. 中华糖尿病杂志, 2014, 10 (22):

参考文献

- 1 孟庆跃. 医疗保险支付方式改革对费用控制的影响分析 [J]. 卫生经济研究, 2002, 8 (9): 18-21.
- 2 何帆, 申思, 方沅湘, 等. 公立医院开展商业医疗保险直接理赔相关问题探讨 [J]. 协和医学杂志, 2014, 3 (2): 246-248.
- 3 王旭鹏. 论中国商业健康保险直付理赔模式的实现 [D]. 成都: 西南财经大学, 2014.
- 4 彭明毅, 左秀然, 杨国良. “互联网+健康商保”在线直赔应用模式实践 [J]. 中国数字医学, 2017, 10 (12): 7-9.
- 5 晓凡. 从“平安”到“健康”医保商保有变化 [J]. 就业与保障, 2007, 3 (4): 28.
- 6 石捷, 毛藤熹, 尹泽. 我国网上保险的现状 [J]. 时代金融, 2013, 12 (36): 4-5.
- 7 邵萍. 乐享寿险理赔“3G”时代 [J]. 今日财富, 2013, 3 (4): 65.
- 8 韩璐. 平安的“三网”合一新型医疗 [J]. 二十一世纪商业评论, 2015, 6 (7): 41.
- 9 左良斌. 与商保公司合作开展医疗管理的实践与思考 [J]. 中国医疗保险, 2014, 3 (4): 64-66.
- 10 沈焕根. 发挥商保补充作用的思考与建议 [J]. 中国医疗保险, 2014, 8 (7): 64-66.
- 11 张苗. 医疗保险 选择商保做补充 [J]. 中国社会保障, 2004, 7 (8): 44-45.

865-869.

- 6 吴云霞, 刘忠军. 对采用新技术治疗寰枢椎疾病治疗费用变化的分析 [J]. 中华医学杂志, 2016, 96 (3): 199-202.
- 7 中国经济网. 1978-2013 中国居民消费价格指数 (CPI) [EB/OL]. [2018-08-31]. <https://wenku.baidu.com/view/dcd9264ceefdc8d377ee3270>.
- 8 国家统计局. 1999-2017 中国居民消费价格指数 (CPI) [EB/OL]. [2018-08-31]. <http://www.stats.gov.cn/tjsj/nds/>.
- 9 原中华人民共和国卫生部. 中国卫生统计年鉴 2014 [M]. 北京: 中国协和医科大学出版社, 2014: 146-149.
- 10 吴云霞, 刘忠军, 张振会, 等. 大数据在临床医学研究中的应用 [EB/OL]. [2018-08-31]. <http://cdas.cda.cn/2016/index.html#subnav1>.
- 11 Jeremy Ginsberg, Matthew H Mohebbil, Rajan S Patel, et al. Detecting Influenza Epidemics Using Search Engine Query Data [J]. Nature, 2009, 457 (7232): 1012-1014.