

# 健康医疗大数据驱动下的精准医疗实施进展\*

黄小龙 罗旭 汪鹏 吴昊

(第三军医大学西南医院 重庆 400038)

〔摘要〕 梳理健康医疗大数据和精准医疗有关概念与特征,探讨健康医疗大数据与精准医疗关系与作用,简要阐述当前存在的问题与挑战并提出发展建议,为推动健康医疗大数据在精准医疗领域的转化应用提供借鉴和参考。

〔关键词〕 健康医疗;大数据;精准医疗

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2017.09.003

**Implementation Progress of Precision Medicine Based on Healthcare Big Data** HUANG Xiao-long, LUO Xu, WANG Peng, WU Hao, Southwest Hospital, Third Military Medical University, Chongqing 400038, China

〔Abstract〕 The paper reviews the concepts and characteristics of healthcare big data and precision medicine, discusses the relationship and function of healthcare big data and precision medicine, and briefly expounds existing problems and challenges, put forward the development suggestions. It could provide a reference for the transformation and application of healthcare big data in the field of precision medicine.

〔Keywords〕 Healthcare; Big data; Precision medicine

## 1 引言

随着信息科技的快速发展,现代社会正加速步入大数据时代,继零售、通讯和互联网行业之后,一场以大数据为驱动的颠覆医疗变革即将到来。健康医疗大数据的开发利用是大数据应用的重要领域,而精准医疗作为未来医学发展的主要趋势之

一,高度依赖于健康医疗大数据的驱动。但目前该领域的研究和应用还存在一系列机制和技术瓶颈,以健康医疗大数据为驱动的精准医疗实施路径仍不明晰。

## 2 健康医疗大数据相关概念

### 2.1 大数据

2008年《自然》(Nature)杂志出版专刊《大数据》(Big Data),“大数据”一词开始正式出现在科学研究领域,迅速成为各领域的研究热点。目前,业界对于大数据概念还没有统一、确切的定义。知名战略咨询公司麦肯锡(McKinsey)将大数据定义为:无法在一定时间内用传统数据库软件工

〔修回日期〕 2017-09-09

〔作者简介〕 黄小龙,硕士研究生;通信作者:吴昊,教授。

〔基金项目〕 国家重点研发计划(项目编号:2016YFB1001404);重庆市社会民生科技創新专项(项目编号:cstc2015shmszx120024)。

具对内容进行抓取、管理和处理的数据集合<sup>[1]</sup>。国际数据公司将大数据定义为：一种新一代的技术和架构，具备高效的捕捉、发现和分析能力，能够经济地从类型繁杂、数量庞大的数据中挖掘出色价值<sup>[2]</sup>。李国杰等<sup>[1]</sup>研究认为大数据是指无法在可容忍的时间内用传统 IT 技术和硬件工具对其进行感知、获取、管理、处理和服务的数据集合。

## 2.2 健康医疗大数据

剑桥大学在《医疗大数据：数据科学的探索与机遇》一文中将健康医疗大数据定义为：主要包括人们日常的健康身体数据、体检数据、病例数据、处方数据、用药情况数据、基因数据、公共卫生范畴内的数据等围绕人体各项健康指标以及与健康行为相关的数据<sup>[3]</sup>。IBM 公司在《大数据医疗应用分析》指南中将健康医疗大数据描述为：在医疗卫生行业中产生的数据，其类型和规模正以前所未有的速度高速增长，以至于无法利用目前主流的软件工具，在合理的时间内撷取、管理并整合成为能够为医疗服务提供更积极高效的决策信息<sup>[4]</sup>。2016 年 6 月国务院办公厅发布关于促进和规范健康医疗大数据应用发展的指导意见，意见重点提及 4 类大数据，包括卫生行业治理大数据、临床与科研大数据、公共卫生大数据和健康管理大数据<sup>[5]</sup>。

## 3 精准医疗相关概念

### 3.1 内涵

2015 年 1 月美国总统奥巴马在国情咨文演讲中宣布启动“精准医疗”计划，大力推动精准医疗系统研究与开发应用。根据美国国立卫生研究院定义，精准医疗是一种充分考量患者在基因、环境及生活方式中存在的个体差异以达成最有效的疾病治疗和预防的医学模式<sup>[6]</sup>。国内部分专家对精准医疗定义达成基本共识，认为精准医疗是集合现代科技手段与传统医学方法，科学认知人体机能和疾病本质，以最有效、最安全、最经济的医疗服务获取个体和社会健康效益最大化的新型医学范畴<sup>[7]</sup>。2015 年 2 月习近平总书记批示科技部和国家卫生计生

委，要求成立中国精准医疗战略专家组，由 19 位专家组成国家精准医疗战略委员会，国内有关研究工作开始大规模布局。

### 3.2 特点

在医疗活动中医生个人的知识和经验往往都是有限的，2004 年世界卫生组织报告指出，全球有 3.5% 至 16.6% 的住院患者接受过不合理治疗，除医生责任心和医疗水平的限制，诊疗模式本身的局限也是显而易见的。精准医疗是一种全新的诊疗策略，需要临床诊疗过程能够更深入、更准确、更全面地捕捉患者个体及其疾病的本质特征，直接“定位”疾病准确的宏观与微观靶点，进而采取精准的诊断决策和治理措施<sup>[8]</sup>。健康医疗大数据则为这种全新的诊疗模式提供支撑，使医生能够实时链接全球最新的知识库和数据库，链接患者从基因谱到症状族的全维度健康信息，获得精准的诊疗提示与帮助，有效减少出现误诊、漏诊和治疗差错可能。

## 4 健康医疗大数据与精准医疗的关系及其作用

### 4.1 关系

一方面，随着医疗信息化建设水平的不断提升，海量基因检测数据、疾病诊断数据（检验、影像及电子病历）等在内的健康医疗大数据类型与规模正在以前所未有的速度增长，呈现出大数据的特征<sup>[9]</sup>。另一方面，随着生物科技与信息技术的不断发展，以大数据、人工智能、生物科技为标志的新技术革命正在引发一场全新的医学模式变革，传统医学模式正加速向 5P 医学模式转变，呈现出预防性、预测性、个体化、参与性和精准性的“5P”医学特性。

### 4.2 数据驱动的医疗服务领域、技术基础和管理规范发生变化

在传统医学模式下，“数据、信息、知识”都只是在医生大脑里汇集、提取、挖掘，而今天日益增长的海量健康医疗信息，已逐步超越医生个人能够全面获得、接收和有效处理的边界，数据驱动时

代的医疗服务领域,技术基础和管理范式都在发生深刻变化。精准医疗突破传统医学模式根本之处,在于对患者从基因图谱到生活习惯等全面信息数据充分掌握和分析基础上,实施对个体有针对性精确医学处置,“精准”基础是“海量”,健康医疗大数据有效应用构成精准医疗基础。正如 Collins<sup>[10]</sup>于 2015 年 2 月在《新英格兰医学杂志》上发表评述所指出,精准医疗未来长远目标是建立一个完善的数据知识体系,聚焦个体化预防和治疗策略,实施更为精确个性化治疗。

## 5 国内外研究现状与发展趋势

### 5.1 健康医疗大数据来源与分类

健康医疗大数据来源分类目前尚无统一标准,综合现有研究发现,健康医疗大数据一般可分为 5 类数据。(1) 医院病历资料为主的医学诊疗数据。(2) 医学文献和指南为核心的医学知识数据。(3) 个体和群体的基因组学数据。(4) 涵盖职业环境和个人行为习惯的暴露组学数据。(5) 以新型物联网和移动监测技术为支撑的个人健康检测数据等<sup>[11]</sup>。健康医疗大数据除具备大数据所具有“4V”特征外,还具有多态性、不完整性、时间性及冗余性等医疗领域特有特征。其中多态性指医疗大数据包括图像、声音、文字等多种形态数据;不完整性指大量数据主要通过人工进行记录,容易形成信息记录的误差和遗漏;时间性指患者就诊、疾病的发病过程及其相关检测诊疗记录具有一定的时序性;冗余性指每日产生的大量数据中可能会包含重复、无关甚至相互矛盾记录。

### 5.2 健康医疗大数据研发与应用

目前国内外健康医疗大数据的应用主要有以下相关领域:(1) 疾病的预测。谷歌公司对 5 000 万条美国人日常搜索频率最高关键词进行分析处理,用来对冬季流行性感冒进行及时而准确预测。(2) 临床决策支持。甲骨文公司通过对电子病历数据挖掘,为临床决策提供指导,包括对临床路径管理提供支持。(3) 临床治疗指南。Chawla 等<sup>[12]</sup> 2013 年

研究建立患者临床症状交流和诊治经验分享平台,通过数据统计挖掘帮助临床治疗指南的制定,寻找特定病人最佳治疗途径。(4) 个性化健康管理。通过健康管理云平台采集和存储个人健康信息,形成个人健康档案,并制定具有个性化特点健康管理策略。(5) 搭建远程医疗系统。借助物联网和移动互联网等创新科技,对患者生理状况实施远程实时监测,为远程诊疗干预提供基础信息和决策依据。(6) 健康危险因素和病因寻找。系统收集患者健康相关危险因素信息,通过大数据分析技术对相关危险因素进行比对研究,以提示和预警疾病风险。

### 5.3 以健康医疗大数据为支撑的精准医疗研究

2011 年日本大阪大学在数百万个单核苷酸多态性公共数据库建立的基础上,通过稳定成熟基因测序分析技术以及大数据分析工具,发现 43 种与高血压相关的易感基因<sup>[13]</sup>。Persu 等<sup>[14]</sup>进一步发现 STK39 基因变异对高血压影响,这一研究结果已经得到几项独立实验的证实。Relling 等<sup>[15]</sup>认为将药物基因组学进行临床实施关键因素应用包括对药物基因组学测试结果的标准化解读,基于测试结果进行处方的临床指引。吴孟超等<sup>[16]</sup>探索肿瘤精准细胞免疫治疗,通过高通量基因测序及医疗大数据分析,获得针对癌细胞特异性新抗原和具有高效应的精准 T 细胞 (PNA-T),富集 PNA-T 对肿瘤患者进行治疗。清华大学、四川大学、北京协和医院等均设立基于数据库和样本库的精准医疗研究机构,研究疾病发生和发展内在规律。阿里云等正筹建以中国人群基因组数据为核心的数据云,目前已纳入超过 40 万份基因组学数据,其深入挖掘将极大提高中国人遗传病诊断效率<sup>[17]</sup>。

## 6 存在问题与挑战

### 6.1 概述

当前,尽管健康医疗大数据应用取得一些探索进展,但仍然没有在日常一线医疗服务中得以广泛应用。事实上,大数据分析技术在精准医疗领域应用目前绝大部分都集中在基于基因组学疾病基因位

点研究及其处理措施（如致病基因敲除、基因修饰）等纯医学技术层面，而这仅仅是大数据推动精准医疗实践内涵很小一部分<sup>[17]</sup>。

## 6.2 缺乏高效统一的健康医疗大数据临床应用知识库与数据仓管理范式

健康医疗大数据存在体量巨大、异构性强、多态性、不完整性和时间序列性等重要特点，而针对这些特点宏观管理在政策法规、共享机制、融合规则等方面顶层设计严重不足<sup>[18]</sup>。没有标准化设计的知识库和数据仓模式，海量健康医疗数据难以产生有用信息和知识，后续应用也就无从推进。如何对复杂的、海量的医疗大数据资源进行有效管理，建立基于大数据的精准医疗知识体系，构建健康医疗大数据应用生态系统，是健康医疗大数据真正应用到临床精准医疗所面临的重大挑战。

## 6.3 驱动精准医疗大数据应用存在系列管理机制、运作模式和技术处理瓶颈

仅靠建立健康医疗大数据知识库，不能直接带来精准医疗实现，必须解决从数据库到临床诊疗行为的流程衔接和模式转化问题，形成各类大数据信息诊疗采纳规则<sup>[19]</sup>，包括：建立一个能够被医生和患者所接受的规范可靠大数据集成分析系统；将描述性的信息呈现为决策辅助要素；建立医生、患者和计算机体系间的医疗行为闭环等。这其中面临一系列软件、硬件和机制问题，必须建立高度集成实用化范式，才能促使大数据精准医疗应用落地。

## 6.4 尚未形成有示范性的以健康医疗大数据为驱动精准医疗服务模式

疾病临床诊疗是医院最核心任务，但由于医疗工作复杂性，大数据真正用于临床精准医疗还没有成熟范例<sup>[20]</sup>。以健康医疗大数据为驱动的精准医疗服务模式需要医生改变传统诊疗习惯，从关注宏观个体人转向微观“数据人”，从相对宏观症状导向疾病诊疗转向以大数据、基因、精准为标志的个性化诊疗。这一转变需要进行系列机制、流程上规范与重构，患者的诊疗参与方式也需要进一步革新，

然而目前这些探索都才刚刚起步，有待进一步研究。

## 7 对策与建议

### 7.1 分析应用影响因素，加强管理顶层设计

当前，健康医疗大数据在精准医疗领域应用水平仍然不高，缺乏便捷有效大数据应用手段，医疗本身复杂性也使其应用进展缓慢。这里面涉及影响因素非常广泛，既有数据本身体量大、类型复杂、存储分散等基础质量问题，又有信息化水平、管理统筹力度、大数据研发能力等管理方面问题，也有集成应用方式、数据挖掘与分析技术等应用推广方面问题。只有找出制约健康医疗大数据应用的关键因素，才能有针对性地进行改进与提高，形成高效健康医疗大数据管理组织构架。

### 7.2 破解技术发展难题，满足医疗实际需求

医院既是健康医疗大数据产生的主要源头，也是精准医疗实施的枢纽与末端。健康医疗大数据辅助精准医疗实现，就是要通过建立医院大数据研发应用软件和硬件平台，为医院开展大数据应用提供存储空间、计算能力、数据分析方法和系列与临床工作平台深度融合的实用化工具。在对健康医疗数据进行采集、存储、清洗、语义分析、后结构化处理的基础上，对诊疗过程具有重要参考价值的用药知识、医学术语、病种诊疗知识库、临床指南和医学文献等系列数据进行多元化的采集和处理，通过这些数据支持来满足精准医疗服务的需求。

### 7.3 建立应用示范模式，促进资源开放共享

现阶段健康医疗大数据在精准医疗领域的应用普及程度不高的原因除了管理研发与应用需求脱节之外，还有一个非常重要原因是缺乏大数据管理与应用学习和参考范式。目前行业中绝大部分健康医疗大数据应用都是依托科研院所或科技公司在进行应用开发，这其中涉及商业利益冲突或者核心技术保护等问题，造成健康医疗大数据资源碎片化严重，开放与共享进程困难重重。必须打破这种分

散、各自为营状态, 构建出互联互通的医院大数据管理与应用范式, 才能有效推进健康医疗大数据应用普及。

## 8 结语

在大数据时代背景下, 医疗思维与技术的颠覆已经在所难免。精准医疗的理念和导向加快了海量医学数据资源的形成, 而海量医学数据资源分析利用又成为实现精准医疗目标关键所在。人们对于精准医疗追求, 催生海量医学数据资源需求, 同时也对健康医疗领域大数据的分析与应用提出更高的标准和要求。尽管目前大数据分析技术在精准医疗领域应用还处于探索起步阶段, 存在管理统筹不足、数据标准化程度低、集成分析技术手段缺乏、数据资源共享不足以及应用推广力度不够等问题, 但随着有关研究不断深入, 健康医疗大数据的数据基础质量将会越来越规范, 集成分析技术将会越来越高效, 管理与研发模式将会越来越系统, 基于健康医疗大数据的精准医疗应用也将会越来越丰富、准确和智能, 健康医疗大数据驱动下的精准医疗将会取得跨越式新发展。

## 参考文献

- 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考 [J]. 中国科学院院刊, 2012, 27 (6): 647-657.
- 金珺, 陈俊滢, 张郑熠. 现有制造型企业基于大数据的商业模式创新: 以中易和为例 [J]. 西安电子科技大学学报 (社会科学版), 2015, (2): 16-23.
- University of Cambridge. Big Data in Medicine: exemplars and opportunities in data science [R]. UK: Cambridge Institute, 2015.
- Kayyali B, Knott D. The Big - data Revolution in US Health Care: accelerating value and innovation [R]. US: Mc Kinsey & company, 2013.
- 俞国培, 包小源, 黄新霆, 等. 医疗健康大数据的种类、性质及有关问题 [J]. 医学信息学杂志, 2014, 35 (6): 9-12.
- National Institutes of Health. About the Precision Medicine Initiative Cohort Program [R]. NIH Online Report, 2015.
- Zheng R. National Estimates of Cancer Prevalence in China [J]. Cancer letter, 2016, 370 (1): 33-38.
- Alyass A. From Big Data Analysis to Personalized Medicine for All: challenges and opportunities [J]. BMC Med Genomics, 2015, 33 (8): 1-12.
- Murdoch TB. The Inevitable Application of Big Data to Health Care [J]. JAMA, 2013, 309 (13): 1351-1352.
- Collins FS, Varmus H. A New Initiative on Precision Medicine [J]. N Engl J Med, 2015, 372 (9): 793-795.
- Cyranoski D. China Embraces Precision Medicine on a Massive Scale [J]. Nature, 2016, 529 (7584): 9-10.
- Chawla NV. Bringing Big Data to Personalized Healthcare: a patient - centered framework [J]. J Gen Intern Med, 2013, 28 (3): S660-S665.
- 吴松. 精准医疗导论 [M]. 广州: 中山大学出版社, 2015: 34-39.
- Persu A. STK39 and WNK1 are Potential Hypertension Susceptibility Genes in the Belhyphen Cohort [J]. Medicine, 2016, 95 (15): e2698.
- Relling MV, Evans WE. Pharmacogenomics in the clinic [J]. Nature, 2015, 526 (7573): 343-350.
- 钱其军, 吴孟超. 肿瘤精准细胞免疫治疗: 梦想照进现实 [J]. 中国肿瘤生物治疗杂志, 2015, 22 (2): 151-158.
- 黄小龙, 罗旭, 汪鹏, 等. 基于健康医疗大数据的精准诊疗实施路径探讨 [J]. 中华医院管理杂志, 2017, 33 (5): 369-372.
- 代涛. 健康医疗大数据发展应用的思考 [J]. 医学信息学杂志, 2016, 37 (2): 2-8.
- 罗旭, 刘友江. 医疗大数据研究现状及其临床应用 [J]. 医学信息学杂志, 2015, 36 (5): 10-14.
- 李维, 陈祁, 张晨, 等. 基于大数据技术的临床数据中心与智能分析应用平台构建 [J]. 医学信息学杂志, 2014, 35 (6): 13-17.