

## ·专家笔谈·

## 再谈大数据时代的小儿外科临床多中心研究

王 培 冯杰雄

**【摘要】** 医疗大数据的挖掘和应用一直是推动临床多中心研究不断向前发展的强劲动力。当前社会背景下,以计算机技术为依托的医院信息化的推广为多中心研究的开展提供了强有力的手段。大数据时代的临床多中心研究不仅有助于更深入地了解疾病,同时也有助于引导个性化诊疗和智能临床决策。笔者旨在结合近年来医疗大数据挖掘和应用领域的新发展动向,对小儿外科的临床多中心研究思路进行更为深入的探讨,以进一步提升我国小儿外科的临床科研水平。

**【关键词】** 小儿外科;大数据;多中心研究

**【中图分类号】** R726 R181.2

**Clinical multi-center researches during the era of big data.** Wang Pei, Feng Jiexiong. Department of Pediatric Surgery, Affiliated Tongji Hospital, Tongji Medical College, Huazhong University of Science & Technology, Wuhan 430030, China.

**【Abstract】** Mining and applying medical big data has been a powerful driver for continuously developing clinical multi-center researches. In current social context, promoting computer-based hospital information technology provides a powerful mode of developing multi-center researches. Clinical multi-center researches in the era of big data not only help us understand human disease more deeply, but also guide personalized diagnosis and intelligent decision-making. This review is intended to further explore the clinical multi-center researches of pediatric surgery in light of new development trends in medical data mining and applications and further boost the clinical research level of pediatric surgery in China.

**【Key words】** Pediatric Surgery; Big Data; Multicenter Studies

随着信息技术的飞速发展及社会经济文化水平的不断提高,医疗大数据时代已悄然而至。几年前,笔者曾就大数据时代下先天性巨结肠多中心研究的趋势进行过评述<sup>[1]</sup>。而近些年笔者发现,面对如潮水般涌来的海量医学数据,如何更高效地对其进行挖掘和利用成为生物医学和信息技术领域共同面临的挑战。故笔者在原有研究的基础上,结合近年来医疗大数据挖掘和应用领域的新发展动向进行更为深入的探讨,以进一步提升我国小儿外科的临床科研水平。

大数据这一概念的提出可追溯到多年前,其本义为海量的、复杂的、可链接的信息<sup>[2]</sup>。医疗大数据包括患者在治疗和诊治过程中产生的基本数据、诊疗数据、医学影像数据、医疗仪器数据、医学管理数据<sup>[3]</sup>,具体可分为:①基因序列、蛋白组学等生物

信息数据;②以电子健康档案、电子病历、医学影像、检验检查等为主的医院医疗大数据;③自我量化大数据;④基于大量人群的医学研究或疾病监测大数据;⑤区域卫生服务平台大数据;⑥网络大数据等<sup>[4,5]</sup>;其基本特征可总结为“4V”(volume, variety, velocity, value)。其中 volume 表示数据容量,医疗大数据的数据容量常常在 PB(1 PB = 250 B)级以上;variety 表示数据种类,医疗大数据常具有不同的数据类型(结构化、半结构化和非结构化)和数据来源;velocity 表示产生和更新速度,代表医疗大数据时效性要求高;value 表示科学价值,尽管医疗大数据利用密度相对较低,但对科学研究的方向具有重要的预测价值。

大数据时代的到来使医学研究发生了质的飞跃,《自然》《科学》《柳叶刀》等杂志相继发表过有关医疗大数据的评述<sup>[6-10]</sup>。不同的研究者一致认为,大数据分析是对随机对照研究的强化和扩展,通过深入了解疾病的起因和结果,为精准医疗提供更好的药物靶点;同时,大数据分析还可广泛应用

DOI:10.3969/j.issn.1671-6353.2019.01.002

作者单位:华中科技大学同济医学院附属同济医院小儿外科(湖北省武汉市,430030),Email:fengjiexiong@126.com

于疾病的监测和预防工作,对改善公众的健康状况具有重大意义。

多年前有学者曾发现染色体的异常可影响肿瘤生长,但并不清楚肿瘤发展与染色体异常的因果关系<sup>[10]</sup>。但在医疗大数据的时代背景下,现代肿瘤学的研究成果得以充分积累,目前已经建立了包含20多种肿瘤、8 207个组织样本、120万个突变位点的大数据,并利用生物信息学的方法筛选出70个原癌基因和50个抑癌基因。单个病人肿瘤基因组和正常基因组的随机组合可产生约1TB大小的数据,而100万个基因组将生成约1EB大小的数据,利用传统模式对这些数据进行存储和分析每年的花费可能高达1亿美元!然而,Bionimbus的开发很好地解决了这一问题,该程序是一个基于云的开源平台,用于共享和分析来自癌症基因组图谱的基因组数据。Megan Mc Nerney等曾使用Bionimbus来追踪一种与急性髓系白血病有关的基因,通过平台分析将涉及到的基因范围缩小到15~20个。Megan Mc Nerney从数据库中挑选了23名患者进行RNA序列的比较,发现这些患者均存在CUX1基因的缺失;动物试验结果也表明,敲除CUX1基因会导致白血病的发生,从而为靶向药物的研发奠定了基础。1854年,霍乱席卷伦敦,流行病学之父John Snow经过长时间的现场调研及后期分析研究才找出霍乱的爆发源头;而今,只需应用全球定位系统信息对流行病学数据进行收集,通过数小时的整理和分析即可解决病因研究的问题<sup>[2]</sup>。

此外,当预期结果为小概率事件时,结合不同地区的研究结果数据进行分析对结论外推能力的提高具有重要意义。如泛北欧国家(丹麦、芬兰、冰岛、挪威和瑞典)曾针对产前暴露于抗抑郁药物对后代的不良反应开展过一项多中心研究<sup>[11-13]</sup>。1995年以来,北欧关节成形术注册协会收集了约100万例原发性髋关节与膝关节置换的病例,解决了单个国家样本量不足的难题<sup>[14-16]</sup>。OHDSI (Observational Health Data Sciences and Informatics)是一个世界性的公益型非盈利研究联盟,该组织中有超过120名研究人员专门针对临床大数据进行分析,目前已利用该网络中4个国家的2.5亿患者疾病数据对各类流行性较高的慢性疾病进行了研究<sup>[17]</sup>。

同时,许多国际组织(如EU-ADR、Sentinel、OMOP、PROTECT和VAESCO)也开始结合多个医疗保健数据库的资料对药物和疫苗安全进行评价<sup>[18]</sup>。EU-ADR于2008年启动了一个基于EHR和生物医

学数据分析的计算机集成框架项目,其中的生物医学数据最初来源于4个欧洲国家(丹麦、意大利、荷兰和英国)的管理数据库和全科医生数据库<sup>[19]</sup>。Sentinel计划成立于2008年<sup>[18]</sup>,是一个由FDA建立的上市后检测系统,利用EHRs(电子医疗记录)前瞻性地监测已上市医疗产品的安全性。OMOP由美国国立卫生研究院基金会管理,旨在确定一个积极的药物安全监测系统,以最大限度地提高药物制剂的效益并减少其临床应用的风险<sup>[20]</sup>。EC资助项目则结合了多个医疗保健数据库提供的信息来评估特定的安全问题(如非甾体抗炎药物化胃肠和心血管风险、药物的心律失常风险、选择血糖剂的心脑血管和胰腺安全性、哌甲酯对注意缺陷和多动障碍的长期不良影响、儿童特发性关节炎患者生物制剂的安全性、epoetins的安全性以及与胰岛素类似物相关的癌症风险等)<sup>[8]</sup>。

大数据的挖掘和开发促进了个性化的健康服务,为疾病的发生发展、预防和治疗提供全面、全新的认识,同时也有利于开展个体化医学<sup>[7,21]</sup>。此外,大数据带来的大样本能够解决流行病学研究中的样本量问题,基于大样本的研究可有效减少抽样误差,提高结论的可信度,而标准化的采集途径能够减少信息偏倚<sup>[22]</sup>。此外,医疗大数据还有助于诊断准确性的提高和患者治疗方案的优化,避免重复诊治,减少住院时间,缩短治疗疗程,提高疗效<sup>[23]</sup>。同时,大数据的应用也改变了临床多中心研究方法,电子病历就是大数据在医学领域运用的一个重要体现,它不仅有效减少了人力和物力的花费,研究的实施过程及结果表述也变得更加简单。此外,特殊标志物的应用可减小试验规模,节约成本。研究者可随时根据研究进度调整和优化试验设计,并根据已有的数据预测受试者可能获得的受益。

临床随机对照试验的结果一直是医学研究公认的金标准。但随着大数据时代的到来,多中心临床研究开始受到越来越多的关注。多中心临床研究是指由多个研究中心的临床医生或科研人员遵循同样的研究设计及研究目的协同完成的临床研究工作<sup>[24]</sup>。由于多中心临床研究要求多个研究中心同时参与,因此可在较短的时间内遴选出足量符合研究要求的病例;相对于单中心研究,多中心研究能充分发挥学术力量的优势,其研究结论具有可信度高、可靠性强等优点。由此可见,多中心临床研究实现了多中心、多学科对同一临床问题的广泛协作研究,对于发挥临床医生的学术优势、促进医

学学科交叉的发展具有重要意义<sup>[25]</sup>。值得一提的是,多中心研究产生的大数据需要进行及时分析和处理。英国、德国等药品检查机构通过分析病人的治疗数据(包括电子病历、医学影像数据、病理报告、门诊处方等)对比了多种干预措施的临床效果,有效地找到了最佳治疗途径,并已获得一定成果。

医疗大数据的优势虽然显而易见,但在实际应用过程中仍存在诸多挑战。如电子病历虽然可以高效地记录病情的动态变化、实验室检测结果和结局,但不能记录发生在医院以外且与疾病相关的事件。另外,不同单位的电子病历数据库往往格式不同,通用性受限问题也是阻碍多中心研究开展的重要因素。许多国家的研究人员和管理人员一直致力于将来自不同来源的医疗数据进行标准化处理,但此项工作困难重重。公共数据模型(CDM)要求每个研究中心根据CDM制定的相关标准输入数据集,该规范规定数据集的类型、结构、变量名、属性以及派生变量的定义,为解决不同研究中心间数据结构和编码系统的差异提供了新思路<sup>[27]</sup>。但该操作也带来了技术、逻辑、伦理和法律方面的一系列挑战(如数据采集和存储的过程中可能因访问不当造成信息泄露,或数据分析与发布的过程中可能因匿名化程度不够而暴露患者隐私等)<sup>[18]</sup>。另外,当诊疗信息转变成科研数据后,医疗大数据“所有权”的归属问题也存在较大争议,这关系到数据使用者、搜集者、生产者三者之间利益协调问题<sup>[28]</sup>。

总之,大数据的时代背景为多中心临床研究的开展提供了良好条件<sup>[29]</sup>。小儿外科医生应充分利用大数据挖掘的科研思维,更加全面地开展小儿外科临床研究。

## 参考文献

- 冯杰雄. 利用医疗大数据开展先天性巨结肠临床多中心研究[J]. 中华小儿外科杂志, 2016, 37(4): 241-243. DOI:10.3760/cma.j.issn.0253-3006.2016.04.001.  
Feng JX. Utilizing big data for conducting multi-center clinical researches in congenital mega-colon[J]. Chin J Pediatric Surg, 2016, (4): 241-243. DOI:10.3760/cma.j.issn.0253-3006.2016.04.001.
- Jee K, Kim GH. Potentiality of big data in the medical sector: focus on how to reshape the healthcare system[J]. Health Inform Res, 2013, 19(2): 79-85. DOI:10.4258/hir.2013.19.2.79.
- 高玮, 傅春瑜, 王剑云, 等. 医疗大数据的统一存储架构设计[J]. 中国数字医学, 2017, 12(4): 4-6. DOI:10.3969/j.issn.1673-7571.2017.04.002.
- Gao W, Fu CY, Wang JY, et al. Uniform Storage Architecture Design of Medical Big Data[J]. China Digital Medicine, 2017, 12(4): 4-6. DOI:10.3969/j.issn.1673-7571.2017.04.002.
- 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. DOI:10.3969/j.issn.1673-6036.2014.06.002.  
Yu GP, Bao XY, Huang XT, et al. Medical and Health Big Data: types, characteristics and relevant issues[J]. Journal of Medical Informatics, 2014, 35(6): 9-12. DOI:10.3969/j.issn.1673-6036.2014.06.002.
- 陆易, 黄正行, 俞思伟, 等. 临床医疗大数据研究现状与展望[J]. 医疗卫生装备, 2017, 38(3): 112-115. DOI:10.7687/J.ISSN1003-8868.2017.03.112.  
Lu Y, Huang ZX, Yu SW, et al. Research situation and expectation of clinical healthcare big data[J]. Chinese Medical Equipment Journal, 2017, 38(3): 112-115. DOI:10.7687/J.ISSN1003-8868.2017.03.112.
- Khoury MJ, Ioannidis JP. Medicine Big data meets public health[J]. Science, 2014, 346(6213): 1054-1055. DOI:10.1126/science.aaa2709.
- Psaty BM, Breckenridge AM. Mini-sentinel and regulatory science-big data rendered fit and functional[J]. N Engl J Med, 2014, 370(23): 2165-2167. DOI:10.1056/NEJMpl401664.
- Murdoch TB, Detsky AS. The inevitable application of big data to health care[J]. JAMA, 2013, 309(13): 1351-1352. DOI:10.1001/jama.2013.393.
- Trifirò G, Coloma PM, Rijnbeek PR, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? [J]. J Intern Med, 2014, 275(6): 551-561. DOI:10.1111/joim.12159.
- Broniatowski DA, Paul MJ, Dredze M. Twitter: big data opportunities[J]. Science, 2014, 345(6193): 148. DOI:10.1126/science.345.6193.148-a.
- Savage N. Bioinformatics: big data versus the big C[J]. Nature, 2014, 509(7502): S66-S67. DOI:10.1038/509S66a.
- Kieler H, Artama M, Engeland A, et al. Selective serotonin reuptake inhibitors during pregnancy and risk of persistent pulmonary hypertension in the newborn: population based cohort study from the five Nordic countries[J]. BMJ, 2012, 344(2): d8012. DOI:10.1136/bmj.d8012.
- Furu K, Kieler H, Haglund B, et al. Selective serotonin reuptake inhibitors and venlafaxine in early pregnancy and risk of birth defects: population based cohort study and sib-



- ling design [J]. BMJ, 2015, 350: h1798. DOI: 10.1136/bmj.h1798.
- 14 Stephansson O, Kieler H, Haglund B, et al. Selective serotonin reuptake inhibitors during pregnancy and risk of stillbirth and infant mortality [J]. JAMA, 2013, 309 (1): 48-54. DOI: 10.1001/jama.2012.153812.
  - 15 Havelin LI, Fenstad AM, Salomonsson R, et al. The Nordic Arthroplasty Register Association; a unique collaboration between 3 national hip arthroplasty registries with 280,201 THRs [J]. Acta Orthop, 2009, 80 (4): 393-401. DOI: 10.3109/17453670903039544.
  - 16 Robertsson O, Bizajeva S, Fenstad AM, et al. Knee arthroplasty in Denmark, Norway and Sweden. A pilot study from the Nordic Arthroplasty Register Association [J]. Acta Orthop, 2010, 81 (1): 82-89. DOI: 10.3109/17453671003685442.
  - 17 Havelin LI, Robertsson O, Fenstad AM, et al. A Scandinavian experience of register collaboration; the Nordic Arthroplasty Register Association (NARA) [J]. J Bone Joint Surg Am, 2011, 93 (Suppl 3): 13-19. DOI: 10.2106/JBJS.K.00951.
  - 18 Park RW. Sharing clinical big data while protecting confidentiality and security: Observational Health Data Sciences and Informatics [J]. Healthc Inform Res, 2017, 23 (1): 1-3. DOI: 10.4258/hir.2017.23.1.1.
  - 19 Stang PE, Ryan PB, Racoosin JA, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership [J]. Ann Intern Med, 2010, 153 (9): 600-606. DOI: 10.7326/0003-4819-153-9-201011020-00010.
  - 20 Platt R, Wilson M, Chan KA, et al. The new Sentinel Network-improving the evidence of medical ~ product safety [J]. N Engl J Med, 2009, 361 (7): 645-647. DOI: 10.1056/NEJMp0905338.
  - 21 Coloma PM, Schuemie MJ, Trifiro G, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU ~ ADR Project [J]. Pharmaco epidemiol Drug Saf, 2011, 20 (1): 1-11. DOI: 10.1002/pds.2053.
  - 22 Costa FF. Big data in biomedicine [J]. Drug Discov Today, 2014, 19 (4): 433-440. DOI: 10.1016/j.drudis.2013.10.012.
  - 23 马晓, 孙瑞靖, 宋小明. 浅析医疗大数据对医学研究领域的创新进展 [J]. 内蒙古科技与经济, 2018 (7): 93-94. DOI: CNKI: SUN: NMKJ. 0. 2018-07-047.
  - Ma X, Sun RQ, Song XM. Preliminary discussion on innovative advances of big healthcare data for medical researches [J]. Inner Mongolia Science & Technology and Economy, 2018 (7): 93-94. DOI: CNKI: SUN: NMKJ. 0. 2018-07-047.
  - 24 丁洁. 关于多中心临床研究的思考 [J]. 中华儿科杂志, 2011, 49 (11): 801-803. DOI: 10.3760/cma.j.issn.0578-1310.2011.11.001.
  - Ding J. Thoughts of multi-center clinical researches [J]. Chinese Journal of Pediatrics, 2011, 49 (11): 801-803. DOI: 10.3760/cma.j.issn.0578-1310.2011.11.001.
  - 25 朱立峰, 刘淑君, 陈德华, 等. 多中心临床大数据平台建设及深度应用 [J]. 大数据, 2018, 4 (3): 46-53. DOI: 10.11959/j.issn.2096-0271.2018029.
  - Zhu LF, Liu SJ, Chen DH, et al. Construction and deep application of multi-center clinical big data platform [J]. Big Data, 2018, 4 (3): 46-53. DOI: 10.11959/j.issn.2096-0271.2018029.
  - 26 万霞, 李赞华, 刘建平. 临床研究中的样本量估算: (1) 临床试验 [J]. 中医杂志, 2007, 48 (6): 504-507. DOI: 10.3321/j.issn.1001-1668.2007.06.009.
  - Wan X, Li ZH, Liu JP. Estimating sample size during clinical trials: Part 1: clinical trial [J]. Journal of Traditional Chinese Medicine, 2007, 48 (6): 504-507. DOI: 10.3321/j.issn.1001-1668.2007.06.009.
  - 27 Reardon S. US big data health network launches aspirin study [J]. Nature, 2014, 512 (7512): 18. DOI: 10.1038/512018a.
  - 28 张妮楠, 史华新, 谢琪, 等. 大数据背景下医学数据共享产生的伦理学问题 [J]. 中国中医药信息杂志, 2018, 25 (8): 9-11. DOI: 10.3969/j.issn.1005-5304.2018.08.003.
  - Zhang NN, Shi HX, Xie Q, et al. Ethical issues of sharing medical data under the background of big data [J]. Chinese Journal of Information on Traditional Chinese Medicine, 2018, 25 (8): 9-11. DOI: 10.3969/j.issn.1005-5304.2018.08.003.
  - 29 Angus DC. Fusing randomized trials with big data: the key to self-learning health care systems? [J]. JAMA, 2015, 314 (8): 767-768. DOI: 10.1001/jama.2015.7762.
- (收稿日期: 2018-11-18)

**本文引用格式:** 王培, 冯杰雄. 再谈大数据时代的小儿外科临床多中心研究 [J]. 临床小儿外科杂志, 2019, 18 (1): 5-8. DOI: 10.3969/j.issn.1671-6353.2019.01.002.

**Citing this article as:** Wang P, Feng JX. Clinical multi-center researches during the era of big data [J]. J Clin Ped Sur, 2019, 18 (1): 5-8. DOI: 10.3969/j.issn.1671-6353.2019.01.002.