

DOI: 10.13288/j.11-2166/r.2014.23.001

学术探讨

大数据时代与中医药学术创新

何 伟*

(陕西中医学院基础医学院, 陕西省咸阳市秦都区世纪大道, 712046)

基金项目: 国家重点基础研究发展计划(“973”计划) 资助项目(2013CB532002)

摘要 随着中医药现代化及国际化需求高涨, 以现代科学诠释中医药学理论实质十分迫切, 但受小数据及还原分析方法限制, 要实现中医药学术理论发展与创新仍步履维艰。随着大数据研究理念的提出, 深入到中医药大数据研究领域去比较大数据与中医药学主要特点, 探讨大数据时代背景下的中医药学术创新的可能性, 试图通过转变中医药学研究思维, 明确中医药大数据的来源及途径, 实现中医药大数据的整合与应用, 发现中医药大数据构建中存在的现实问题等, 以促进中医思维意识、中医基础理论及诊疗理论的创新与发展。

关键词 大数据时代; 中医药学; 学术创新

大数据时代极大地提升了我们收集海量数据、获取大量信息、预测事物及现象内在规律的能力, 也迫使现代社会的各个行业都在深刻反思如何去顺应大数据时代发展潮流, 以促进自我适应及发展。回首中医药学理论体系的形成历程, 展望中医药学术发展的未来趋势, 如何保持中医药学旺盛的生命力, 是中医药学界面临的科学难题。受现代科学还原分析研究方法的制约, 中医药学的理论发展及学术创新难以取得实质性进展。借助于大数据研究理念及方法, 围绕大数据研究特征与要求, 结合中医药学术特点, 通过明确中医药大数据来源、构建及分析方法, 阐释中医药大数据研究中存在的现实问题, 将对中医药学术创新与发展具有重要的战略意义。

1 大数据时代带给中医药学术创新与发展的启示

英国数据科学家维克托·迈尔—舍恩伯格阐述了大数据时代处理数据具有三大理念^[1]: 是全体数据, 不是随机抽样; 是混杂性, 不是精确性; 是相关关系, 不是因果关系, 揭示了大数据研究的基本特征及要求。其实针对研究对象整体或整合若干类似独立研究数据以验证研究结论的实用性在临床研究中多有体现, 如循证医学的系统评价/Meta 分析, 通过全面收集并整合单一、相似研究文献, 以

获取更高级别的研究证据。真实世界研究理念强调在临床实际用药情况下的药品安全性、有效性及经济性评价, 即通过“真实世界样本”来反映真实世界总体^[2]。比较效益研究理念注重比较不同临床治疗方案的受益和风险, 从而获得较为明智的临床决策^[3]等, 在研究理念上已有契合大数据研究趋势。相对而言, 大数据时代研究视角更为宽广, 思维方式更为开放, 产生的影响也更为深远。从研究思维、理念、内容及方法间的逻辑关系来看, 研究理念的产生深受研究思维方式的影响, 而研究内容及方法是研究理念的外在表现形式。因此, 大数据时代带给中医药学的最大启示在于, 重新审视并变革我们已有并固化的中医药研究思维, 去洞见中医药学术规律特征, 预测中医药学术未来发展趋势。

2 大数据时代的基本特征与中医药学的主要特点

2.1 全数据模式与整体观

大数据时代强调全数据模式, 即分析全部数据, 而不是随机抽样的部分数据, 以部分去反映整体特征是在数据获取技术及条件限制下的无奈之举。因此, 小数据追求研究精度, 以期提高研究结果的可信性, 而大数据的产生及获取使情况发生了根本转变, 以全部数据反映整体特征才是最佳研究方式, 故大数据研究允许混杂存在, 更为关注研究结果的实用性。

* 通讯方式: hewei7826@163.com, (029) 38185123

整体观贯穿于中医药学理论与临床各个领域,强调结构及功能的统一性及完整性,即把研究对象看作有机整体,反对孤立、片面、局部地阐释事物及现象的内在规律及特征。所以在研究对象、研究角度及层次认识上,中医药学与大数据研究均立足于整体,均以揭示事物的整体规律为指归。

2.2 相关关系与功能观

相关关系的核心是量化两个数据值间的数理关系,可以用强相关关系及弱相关关系表述研究对象间的关联程度,通常我们更习惯于揭示两数据间的强相关关系,而较少去关注弱相关关系。相对于产生相关关系的原因,我们更倾向于获取产生相关关系的结果,因为明确事物间的相关关系,已经足以使我们更清晰地认识事物,更快捷地采取干预措施。

中医药学术理论的形成及发展并不缺乏形体结构研究的实例,但在传统思维方式的影响下,中医药学难以将实证研究方法作为主要研究手段,而是以对立统一、普遍联系的观念去阐释结构基础上所具有的功能,以明确两者间的相关关系。如藏象学说体系的形成就是通过观察外在体、窍、志、液等生理及病理表现,来验证与内在脏腑功能的相关关系。因此,大数据研究与中医药学研究均注重于揭示研究对象间的相关关系。

3 大数据时代实现中医药学术创新的可能性

大数据时代通过大数据分析,以发现其背后隐藏的事物间的相关性,或事物内部的规律及特征。中医药学通过吸纳老子“道”文化思想及内涵,将自然之道比类人体之道,促进了自身整体观念的形成及发展,提倡人类生存应该顺应自然及生命的规律及法则,进而明确以揭示人体生命现象规律为中医药学的根本任务。因此,大数据时代试图预测及发现事物的规律,与中医药学通过反复的临床实践及哲学思辨力图阐释人体生命的基本规律有相通之处。借助大数据研究思维及方法,促进中医药学术发展与创新,具有明显的理念优势。

由于中医学是依“观物取象”和“立象尽意”方法形成的人身科学,主要是意象思维的产物^[4],故在中医药学理论体系的形成及构建过程中,虽有一定的结构解剖及形态观察基础,但主要采用“远取诸物”、“近取诸身”的感悟及体悟方法,通过对自然社会及人体自身生理病理现象的长期观察,对临床实践经验的积累及升华,经验证及修正后逐渐取得共识。所以建立于个体临床实践基础上

的中医药学术理论,历经数千年漫长而坎坷的形成与发展过程,期间虽有部分学术理论创新,但其形成具有一定的被动性、偶然性和不确定性。而在信息爆炸的大数据时代,医疗信息产生的速度以及数量空前迅捷和庞大,尤其是我国医院信息化建设的强化,如电子病历系统的普及应用,使获取临床诊疗大数据成为可能,通过中医药大数据的共享、整合、集成、分析,可以极大地提高中医药学理论研究效率,发现现代中医药学新规律及新特征,最终实现中医药学基础理论、临床诊疗理论创新,中药临床评价方法创新,加快重大中药新药创制进程。

4 中医药学术创新的大数据来源

中医药学术创新的大数据主要来源于几个方面。①临床医疗数据:临床医疗数据是中医药学术创新的主体数据,主要来源于医院信息系统(hospital information system, HIS),如影像存档和通讯系统(picture archiving and communication systems, PACS)、临床信息系统(clinical information system, CIS)、实验室信息系统(laboratory information system, LIS)等,其中临床信息系统的电子病历(electronic medical record, EMR)数据尤为重要。由于我国中医院电子病历包含中医四诊信息数据及实验室检测数据,而西医院电子病历缺少中医舌脉信息,需要整合两者的电子病历系统数据,以做到优势互补。②中医药文献数据:包括期刊文献、会议文献、博士论文、硕士论文、报纸、专利、标准、图书文献(中医古籍、现代医著)等。检索CNKI医药卫生领域文献结果显示:中医学、中药学、中西医结合领域的期刊文献达百万篇以上,国内/国际会议文献达十万篇以上,而随着我国中医药事业发展,中医药科研水平深化,中医药文献数据量将进一步迅猛增长。③临床研究数据:主要包括I-IV期临床试验数据,自发呈报及主动监测安全性数据,中药上市后临床有效性、安全性、经济性研究数据等。④实验研究数据:主要包括动物实验、临床前药理学、毒理学实验等数据。⑤其他数据:如网络信息数据,药品、医疗器械销售信息数据,有助于分析亚健康人群及非医院门诊医疗患者发病规律特征。

5 构建中医药学大数据的关键环节分析

5.1 变革中医药创新研究思维

大数据时代背景下的中医药学术创新首先不是收集大数据本身。大数据不是研究技术方法,只是

实现大应用的前提, 中医药研究思维的变革才是中医药学术创新的关键。如在传统的还原分析思维影响下, 我们普遍认为, 医学实验检测是确定疾病发生的必须手段, 但基于海量网络检索信息数据构建数学模型, 可以准确预测甲型 H1N1 流感暴发流行的时间及区域, 因此, 借助网络信息大数据可以预测疾病发病。所以中医药学创新研究不应仅着眼于动物实验及临床试验, 去苛求研究水平、层次及研究技术的先进性, 也应着眼中医药学术理论形成根源, 利用临床医疗大数据创新中医药学术理论, 并培养从其他现代信息数据中发现中医药学术规律的能力。

5.2 中医药大数据的建设及获取

中医药大数据可以来源于网络信息、电子信息等非中医药数据, 但更重要的是通过中医药本身大数据来总结中医药学理论及演变规律。目前我国中医药大数据除期刊文献数据外, 尚缺少大规模的、规范化的数据库, 虽然实际产生的中医药诊疗数据十分庞杂, 但需通过大范围、多层次的信息化数据库建设, 才能得以保存并应用于大数据分析。据 2007 年我国卫生部信息中心对 3765 所医院临床信息系统应用比例调查结果显示, LIS 系统应用率为 26.43%, PACS 系统应用率仅为 9.00%^[4], 提示我国的临床医疗信息建设及应用仍处于发展的初级阶段。此外, 中医药大数据的获取也非易事, 个体行为难以获取足够的大数据并得到大应用, 信息数据及时有序公开与自上而下的政策引导极为必要。在大数据获取中还存在诸多细节问题, 如中医药术语的规范化应用、患者医疗信息保密及伦理原则等, 这也是影响大数据构建的重要因素。

5.3 中医药大数据的取舍

单一中医药数据难以满足大数据研究对数据量及数据混杂性的需求, 但仍不能认为数据量越大越好, 过大的数据代表过多的处理时间及工作量, 且研究结果的准确性并不一定会随着数据量的无限扩大而得到显著提高, 因此, 数据的取舍是大数据获取及整合时必须认真对待的重要环节。

5.4 中医药大数据的整合与应用

中医药数据系统大多为文本、图像等非结构化数据, 现实世界产生的各种信息也多以非结构化形式存在, 同时又要对不同异构交叉数据进行整合, 为大数据建构及分析带来巨大挑战, 只有多学科技

术结合方法才是整合与分析大数据的有效手段。如数据可视化技术可以直观展示及理解大数据研究的相关关系及规律; 改进的数据挖掘及机器学习技术仍适用于大数据分析; 数据仓库技术包括数据清理、数据集成和联机分析处理, 其中的联机分析处理技术 (online analytical processing, OLAP) 具有汇总、合并和聚集功能, 以及从不同角度观察信息的能力^[5]。由于大数据研究方兴未艾, 尤其是在中医药研究领域, 尚缺乏相应的研究基础及实例, 以海量元数据为基础, 利用数据仓储、数据挖掘等技术, 解决复杂异构数据库间的集成整合, 将有利于大数据研究的开展, 而大数据的交叉、整合及分析方法仍需实践探索与验证。

6 结语

随着人们处理数据思维模式的转变, 收集、存储、传输、分析数据能力的提升, 面对庞大而繁杂的信息数据, 我们能够站在前所未有的研究高度, 以整体分析的宽广视角去挖掘并预测事物及现象的整体规律, 切实提高工作及生活效率。而对于中医药学术创新来说, 大数据分析虽可满足中医药学持续发展的需求, 但中医药大数据建设的开展范围及程度差异巨大, 中医药大数据的获取、处理及分析过程仍存在各种理念及技术上的不足, 若要真正实现中医药大数据的大应用, 仍需要理论及实践上的不断尝试与探讨。

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: 生活、工作与思维的大变革 [M]. 盛杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2013.
- [2] 任德权. 真实世界研究、GCP 和风险最小化 [J]. 中国中药杂志, 2012, 37(18): 2681-2682.
- [3] 谢雁鸣, 廖星, 王永炎. 将比较效益研究理念引入中医临床评价研究 [J]. 中西医结合学报, 2011, 9(8): 813-818.
- [4] 刘长林. 中医学: 象科学的代表 [J]. 中国中医基础医学杂志, 2004, 10(3): 161-164.
- [5] 中国医院协会信息管理专业委员会. 《中国医院信息化发展研究报告(白皮书)》(摘录二) [J]. 中国数字医学, 2008, 3(7): 7-19.
- [6] 雷国华. 计算机辅助医学技术导论 [M]. 东营: 中国石油大学出版社, 2011: 132.

The Age of Big Data and TCM Academic Innovation

HE Wei

(School of Basic Medical Sciences, Shaanxi University of Chinese Medicine, Shaanxi Province, 712046)

ABSTRACT With the increasing demand of modernization and internationalization of traditional Chinese medicine (TCM), it is urgent to make modern scientific explanations on the essence of TCM theories. However, under the limit of small data and reductive analysis, we were facing great difficulties to achieve academic development and innovation of TCM. With the proposal of the concept of big data, we were obliged to dig in the research field of TCM big data to compare the main characters of big data and TCM, to discuss the possibility of TCM academic innovation under the big data background, to try finding out the origins and means of TCM big data by changing the thinking mode of TCM research, and finally to achieve the integration and application of big data with TCM. We explored the realistic problems existing in the construction of TCM big data, as to promote the innovation and development of TCM thinking, TCM basic theories and TCM diagnosis and treatment theories.

Keywords big data; traditional Chinese medicine; academic innovation

(收稿日期: 2013-11-08; 修回日期: 2014-01-16)

[编辑: 黄健]

中医杂志社医海林音像书店邮购图书目录

●国医大师临床丛书
李济仁医论医选集 98
中医古籍文献学 (第2版) 228
张琪学术思想探赜 78
●名老中医方药心得丛书
祁宝玉眼科方药心得 50
晁恩祥临证方药心得 58
俞长荣临床经验集 68
姜良铎内科方药心得 68
段富津方剂学讲读实录 78
周文泉老年病临证经验集 79
徐福松男科纲目 88
白兆芝临证经验集萃 88
●名医讲学荟萃
裘沛然讲《内经》38
张琪讲临床 42
孟庆云讲中医经典 58
●农村卫生适宜技术推广丛书
地方性疾病中西医诊疗技术 27
中药炮制实用技术 29.8
农村实用针灸技术 34
农村实用液体疗法 37
灾害医学救治技术 39.8
●社区医师中西医诊疗规范丛书
心血管疾病 22
社区康复 22
肿瘤 24
呼吸疾病 24
骨伤病 24

神经精神疾病 24
儿科疾病 25
老年疾病 25
风湿免疫及内分泌疾病 28
社区常见疾病预防与保健 29
社区急救 29
流行病与传染病 29
社区医学检查诊断技术 29.8
肾脏及血液疾病 29.9
耳鼻咽喉及口腔疾病 29.9
泌尿生殖疾病 30
外科疾病 30
眼科疾病 35
消化疾病 35
社区护理 35
●中医药科学方法学研究系列丛书
中医药临证特效实例与方法学思考 50
中医学科学方法特征与沿革 69
当代中医药学家谈科学思维与方法 88
古代中医药名家学术思想与认识论 98
中医药理论技术发展的方法学思考 118
中医临床基础医学研究 128
●重读中医经典丛书
金匱要略临床运用 48

金匱要略临床精要 58
伤寒论临床精要 59
伤寒论临床运用 59
内经临床精要 59
伤寒论临床发挥 68
内经临床运用 79
●其它图书
抗微生物药物合理应用手册 39
让“胰岛”动起来——我是这样战胜糖尿病的 (第3版) 39
中药药对表解 (修订版) 39
手法治疗骨科软组织损伤图释 39.8
章次公《药理学》点校 48
病毒性感染性疾病中医治疗学概要 54
全国名老中医柴瑞霭临床经验集萃 78
病毒性传染病中医治疗概要 58
黄德彰医文精华 68
陈雨苍中医妇科临床经验 68
中医学概论 (英文版) 68
中医学导论 69.8
常见肿瘤的非手术治疗 79
张灿理医论医案纂要 79
中医药学名词内科学妇科学儿科学 88
脊髓损伤的中西医康复治疗 98
中国的藏医 (汉文、藏文、英文) 98

现代中成药的药材炮制 98
刘炳凡医论医案 118
华廷芳学术经验集 128
临床常见疾病影像诊断及治疗原则 128
糖络杂论 (第2版) 128
伤寒论方剂当代研究 128
柴浩然医论医案集 128
林沛湘学术经验集 138
康复评定学 148
现代脉诊学 148
岭南本草集锦 148
章次公医案经验集 (增补版) 158
中药血清药物化学 188
天然产物化学研究 198
广西特色中草药资源选编 498
千金翼方论译 110
当代中医外治精要 48
济阴济阳纲目 140
医学入门 68
奇效良方 77

以上书均免费邮寄,欲购者请汇款至北京市东直门内南小街16号中医杂志社医海林音像书店图书部收,邮政编码:100700。电话:(010)64035632。请在汇款单附言栏写清所购书名、册数及邮编。