

大数据时代下探讨癌症早诊早治资源信息化综合建设的必要性

宋国慧¹, 魏文强²

(1. 磁县肿瘤防治研究所, 磁县肿瘤医院, 河北 磁县 056500; 2. 国家癌症中心/国家肿瘤临床医学研究中心/中国医学科学院北京协和医学院肿瘤医院, 北京 100021)

摘要: 大数据(巨量数据集合)时代的到来是必然性的, 给肿瘤预防及研究带来了新的机遇与挑战。我国肿瘤的相关防治及早诊早治工作的蓬勃发展, 其综合信息化建设及数字网络化是重要趋势。为此, 本文通过以上消化道癌大人群筛查为例, 探讨人群为基础的癌症筛查及多源数据的资源整合及信息化建设的必要性, 不断完善及修订相关资源数据库的框架设计及改进, 适应数据数字化的转型。大数据时代下癌症早诊早治资源整合及信息化建设非常必要, 具有深远的意义。

关键词: 大数据; 癌症防控; 早诊早治; 信息化

中图分类号: R73 文献标识码: A 文章编号: 1004-0242(2019)06-0401-04

doi: 10.11735/j.issn.1004-0242.2019.06.A001

Construction of Comprehensive Information Resource System for Early Diagnosis and Treatment of Cancer in the Big Data Era

SONG Guo-hui¹, WEI Wen-qiang²

(1. Cixian Cancer Prevention and Treatment Institute, Cixian Cancer Hospital, Cixian 056500, China;

2. National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China)

Abstract: The big data era (big data sets in IT term) brings new opportunities and challenges for cancer prevention and research. With the rapid development of cancer prevention and treatment in China, the comprehensive information construction and digital network are important trends. Taking large population-based screening of digestive tract cancer as an example, the current study is to explore the necessity of integration and construction of comprehensive information system based on multi-source data. The framework design and improvement of related databases should be constantly improved and revised to adapt to the transformation of data digitization. In big data era, the integration and comprehensive information construction of early diagnosis and early treatment resources are of great importance, which will deeply influence the cancer prevention and research in the future.

Key words: big data; cancer control; early treatment; information construction

随着互联网、物联网、云计算等信息技术与通信技术的迅猛发展, 数据量的暴涨成了许多行业共同面对的严峻挑战和宝贵机遇。网络数据与技术作为信息科学、社会科学、网络科学和系统科学等相关领域交叉的新兴学科方向, 正逐步成为学术研究的新热点。大数据的深挖掘、大规模利用是新兴产业界的立足点, 即便针对大数据的研究目前还没有建立一套完整的理论体系, 也缺乏高效快速的处理、分析与挖掘的算法与范式, 但大数据的应用前景毋庸置疑,

可为世界经济创造巨大价值, 提高部门的生产力和竞争力, 未来属于那些能够把数据转换成产品的公司和人群^[1]。所谓大数据, 即以“海量”、“巨量”级多类型数据为基础, 通过对其进行迅速地采集、分析和处理, 实现数据应用和管理价值的创造过程。这个过程, 就是将复杂类型的、巨量的碎片化的不可用的信息, 通过收集、分析和处理, 提炼为整合化的可用信息, 实现价值的再创造^[2]。我国肿瘤的相关防治研究及早诊早治工作的蓬勃发展, 其综合信息化及数字网络化是重要趋势, 加强其整体信息规范统一化及共享化建设是非常有必要的。本文以我国人群为基础的上消化道癌筛查为例, 对人群癌症筛查从基

收稿日期: 2019-03-05; 修回日期: 2019-04-10

基金项目: 食管癌专病队列研究项目(2016YFC0901400);

邯郸市科学技术研究与发展计划项目(1823208115ZC)

通信作者: 魏文强, E-mail: weiwq@cicams.ac.cn

础基本信息及危险因素信息搜集、登记、统计、治疗、随访等大量资源信息探索未来前景及综合利用进行讨论,旨在进一步整合早诊早治大数据库链接资源的综合数字化,探讨在大数据时代下癌症早诊早治资源信息化建设的必要性及重要性。

1 人群为基础的癌症筛查数据整合及数字信息化建设势在必行

癌症是威胁人类生命健康的最重大疾病之一。各个国家和地区之间不同的癌症特征表明,地理多样性仍然显著,在社会和经济转型的不同阶段,地区性的风险因素仍然存在。为降低其发病率和死亡率,国家调整公共卫生资源,开展积极有效的早诊早治干预措施,2004年和2005年原卫生部及中国癌症基金会及部分省卫生厅共同建立癌症早诊早治示范基地,并实施推广,癌症早诊早治作为癌症控制的主要策略之一,并提出具体的措施,如研究并制定主要癌症筛查及早诊早治指南、规范,在农村高发区分别建立食管癌、胃癌、肝癌及鼻咽癌的早诊早治示范基地等,并建议建立可持续性研究队列及相应生物样品库,为促进癌症防治研究及资源共享搭建平台等^[3-7]。至今10余年来,专家模式转变政府模式,技术模式得到了广泛推广,涌现出一大批优秀示范现场,尤其在农村癌症早诊早治项目得到最大模式化,截止2018年包括上消化道癌、结直肠癌、肝癌、鼻咽癌和肺癌,共计259个项目点,覆盖全国除台湾、香港、澳门外的31个省、市、自治区,其中项目点上消化道癌为194个、结直肠癌33个、肝癌13个、鼻咽癌8个、肺癌11个,上消化道癌早诊早治的农村模式逐步推广及应用,覆盖的人群逐步增大,并取得良好的社会效益。初期以示范基地为基础逐步实施并推广,建立各领域的数据采集流程,为当时汇总、分析、研究提供了有效的科学资料;但随着项目点数量增加,覆盖人群范围空前扩大,加强落实各环节实施技术及方法管理的同时,相应的基础研究数据资料规范性、统一性、及时性及拓展应用性,在现有的单机模式已经不能满足要求,形成的孤岛数据将成为综合利用和发展的瓶颈。

相关数据整合成为一个棘手问题,单机版数据库采集软件,必须对其进行转换等处理后才能充分

应用^[8]。数据量的不断累积、计算机操作系统的不断更迭升级,有些项目点虽然已经创新落实网络服务器,但在当前大数据时代背景下,其发展仍然缺乏系统性和综合性及可持续性。没有形成整体层面的宏观规划和科研层面及临床层面、随访机制相互配合的系统性对接;大数据背景下,医疗信息发展需要多种技术配合,并进行有机整合,最大限度地实现信息资源的融合共享开发利用,即便在相关数据应用与研究上已有综合性的整体分析及成果,依然缺乏多维度及更深层次的深度挖掘,只是“数据大”也不能称之为“大数据”。癌症防控措施的早诊早治项目涉及众多领域,信息采集量大,在病因探讨等预防预警调查中的调查问卷繁多,并且要求细致,同时,各项目点技术水平参差不齐,还不能实现系统集成的要求,综合信息化要形成国家、省、市、县(区)协同发展的信息网络,需要标准化、规范化。区域医疗信息化是未来医疗行业发展的方向,意义十分深远,它预示着利用信息技术对医疗行业资源进行重新整合的新方向,以信息共享为驱动的医疗服务模式的改变必将为人民的健康提供更加优质、便捷、安全、高效、可及的医疗卫生服务^[9]。在癌症早诊早治信息化管理及防控中,如何将涉及相关零散的孤岛数据迅速地收集、处理、整合,并利用网络技术形成癌症防控资源信息化开辟新的领域,有必要进行一场数字革命,落实其资源数据的信息化是未来发展的必然趋势,信息发展中的大数据投入及应用价值也非常重要。

2 加强基层资源数据标准化及共享服务的利用率

各环节信息采集标准统一性较弱,例如上消化道癌症早诊早治人群筛查中就需要各类数据为基础,主要来自于居民危险因素信息、各项诊疗过程和医疗管理记录信息,胃镜诊断、病理诊断、随访信息、社会信息的对应以及大量的图像资源信息和图像描述信息等。涉及医疗各个领域,这在医疗单位信息采集系统中,虽都进行有关项目内容的采集,但是各自为政,在基础人群资源收集及医疗临床系统中尤为突出,缺乏有效支持综合数据的利用及系统整合。我国尚无统一的巨量级数据收集规范,数据的真实性、准确性得不到保证。一方面在数据收集上缺乏统一

标准,另一方面亟待建立统一的质量标准^[10]。同时大数据时代下信息化发展需要多种技术支持,如数据存储技术、数据分析处理技术、信息融合技术、互联网、移动互联等,这些技术需要相互衔接整合。一方面这些技术本身缺乏统一的标准,这些技术之间,缺乏统一的融合性技术标准,给技术整合造成了一定的障碍。

丰富的高危人群危险因素信息,丰富的地域环境等空间信息,在癌症早诊早治的防控及诊疗过程中均产生了多源及来自各方面的大量信息内容包括个人健康、医疗信息等数据资源。同时,对于癌前病变治疗前、治疗中、治疗后的健康状况、间隔时间、用药情况、饮食结构、心理因素、家族信息、生存质量等积累了丰富资料,而这些数据资源目前均缺乏深入有效的分析,更没有完全实现电子化、信息化、综合化,并在产生一定的决策作用尚未充分体现,转换服务及共享应用率较低。大数据时代下的医疗信息发展,需要以个人健康、医疗信息为基础,近年来我国信息发展重点主要集中在城市,城市各大医院均加大了医疗信息化投入,逐步实现了病历数字化、医学影像数字化和数字化医院建设。基层乡镇医院的信息化投入不足,数字化、网络化基层的重视不够,城乡之间发展不平衡导致了基层数据收集量不足^[8],对标准化及综合建设造成影响。

3 数据资源整合及信息化建设必要性的探讨

3.1 不断完善整体及综合规划

当前大数据时代下,应不断完善及修订相关资源数据库的框架设计及改进,适应数据数字化的转型。首先,开展历史资源信息资源的整合,完善大人群综合数据队列建设,转变思路,拓展数据库的数字模式转换,应完善顶层宏观层面的全面规划,并对已经产生的有关领域信息化发展的大数据应用给予指导;其次,针对大人群筛查实施中其拓展的科研领域进行梳理并结合阶段数据分析结果,进行修订及改进流程,促进综合数据的可持续发展规划,同时也扶持不足的领域,加强综合整体发展;最后,应完善微观层面相关基层单位的扶持和指导,包括策略的制定、规范化制定、标准网络落实等。培育基层骨干人

才,使基层防治队伍不断成熟、壮大、发展。

3.2 加强大数据标准化的转化并深度挖掘拓展综合利用

在统一数据标准化的基础上,提高数据的利用率。既要整理、整合现有的多源数据,又要建立多层次的数据共享平台,通过对单位及个人基本信息数据的分析和处理,使我国癌症早诊早治大人群的大数据在其防控领域及健康管理中发挥更大作用;通过对癌症相关领域的整合,逐步将现有的“数据大”努力转化为“大数据”,为“深度学习”及“人工智能”做好准备,通过不断分析发现问题,纠正问题,以期提高业务技能,提高肿瘤防治决策水平,真正使大数据在疾病管理中发挥更大的作用;基层也通过对积累数据进行研究,真正做出为当地行政部门提供有力的科学信息,利用国家大人群筛查平台,提高利用率,落实转化机制,更好地服务百姓,同时实现基层医院社会效益的不断提高,从而带动基层医疗单位的经济效益及管理水平的提高。

3.3 重视基层信息化建设及数据安全

我国目前基层医疗整合信息化投入不足,数据量的多元化收集量远不如城市。重视基层大人群基础资源收集落实标准化、系统化、规范化管理,首先应加大基层医疗及个人信息化建设,逐步实现基层医院数字化;并不断加大整理、整合力度,依据自身资源优势,加强信息细节数据库管理;拓展综合性基层人才的培养,加强基层癌症防控信息网络、移动互联等能力建设,提高信息化投入,推动区域医疗科研数字化,为相关专家和行业精英搭建一个交流平台,确保人工智能系统广泛应用之前,确保系统能力的可控的、明确的、已充分理解的方式安全操作,在信息化建设中应设计资源综合数据的安全也不容忽视,遵循统筹规划、资源共享、互联互通、安全可靠为原则。信息化是当前发展的大趋势,其资源开发利用也应做好责任管理,转变传统的安全防范观念,转变新时代新思维,不断完善数据信息安全保障措施建设等。

3.4 高危人群筛查资源信息化建设的必要性

大数据早期癌症预测平台的搭建,应用精准预测癌症高风险人群,及时动态预测早期癌症,帮助患者早诊断、早治疗、早康复具有重要意义。相关医务人员要充分学习掌握项目的基本内容、知识原理及

技术应用,积极整合医疗数据资源,组建多学科专家研究团队,开展大数据在癌症等重大疾病防治领域的全方位探索与应用,进一步提高当地医疗卫生水平。所以,很多数据也就仅仅只是数字,无法去转化或为防控建设及社会利益服务,成了一个华丽的摆设或者鸡肋。大数据环境下数据来源非常丰富且数据类型多样,存储和分析挖掘的数据量庞大,对数据拓展展现的要求较高,并且很看重数据处理的高效性和可用性^[3]。高危人群筛查多年来虽然在其采集、处理、存储上做出成绩,随着时间、范围、区域不断增加及信息系统更新升级等其中各类标准均也出现一些差异,不断完善多源异构数据进行转换清洗,拓展其有效利用、分析,并挖掘其中隐含信息,将分散的数据变为有用的信息,加以创新和积累形成一定标准化数据才具有资产价值。所以,高危人群筛查及早诊早治的综合人群预防多源信息化的规范整合建设非常必要。

4 肯定与发展

大数据是一个宽泛的概念,见仁见智。无一例外地都突出了“大”字。诚然“大”是大数据的一个重要特征,但远远不是全部。本文重点讨论,就目前正在开展的癌症早诊早治方面进行建设的大数据综合资源必要性和可行性进行分析,如目前实施的上消化道癌早诊早治已经历14年,涉及188个项目点,搞好信息化建设的危机感和紧迫感,高度重视信息化建设,认真抓好、做好信息建设。需要不断完善信息网络、信息资源、信息技术、信息产业、信息法规、信息环境与信息人才六大要素,向着智能化与网络化发展,脑力与体力相结合,国家与基层相结合,在癌症防控及早诊早治实践中,所涉及的资料量规模巨大,其累计数据的统计、整合与综合利用不足。总之,伴随着一体化的网络大数据时代,早诊早治项目综合信息落实,应进一步提高信息化认识,加强信息化建设步伐已成为必然,不断更新观念,提高对信息化建设的重视度,以政府信息化为主导,以实现综合信息化为基础,实现全社会的信息化、综合化为目标,一定会使我国肿瘤防治整合医学逐步驶入快车道。

大数据的核心能力,是发现规律和预测未来。我国已经迈进肿瘤大数据时代,在不远的将来,全人群

肿瘤发病死亡和生存数据,以及诊疗数据的分析利用势必在肿瘤防控中发挥更大的作用^[11]。如今AI(AlphaGO)+已经成为继互联网+后再次让各医疗领域未来发展的工具,智慧医疗及智能化有关的科研探索等逐步显现,未来智慧筛查指日可待,癌症预防与人工智能的发展,即将颠覆各个行业,其核心是数字医疗领域、科研领域的高质量的综合数据是前提,大数据时代下癌症早诊早治资源整合及信息化建设非常必要,一定具有深远意义。

参考文献:

- [1] Wang YZ, Jin XL, Cheng XQ. Network big data: present and future[J]. Chinese Journal of Computers, 2013, 36(6): 1125-1135.[王元卓,靳小龙,程学旗.网络大数据:现状与展望[J]. 计算机学报, 2013, 36(6): 1125-1135.]
- [2] Wu HQ. Opportunities and challenges in the era of big data[J]. Qiushi, 2013, 4: 47-49.[邬贺铨.大数据时代的机遇与挑战[J]. 求是, 2013, 4: 47-49.]
- [3] Dong ZW, Qiao YL, Li LD, et al. A report of cancer high incidence scene in China[J]. China Cancer, 2009, 18(1): 4-9.[董志伟,乔友林,李连弟,等.中国癌症高发发现场报告[J]. 中国肿瘤, 2009, 18(1): 4-9.]
- [4] Dong ZW, Qiao YL. The practice and discussion of population-based cancer screening program in China [J]. China Cancer, 2009, 18(9): 686-689.[董志伟,乔友林.中国癌症早诊早治的实践与思考 [J]. 中国肿瘤, 2009, 18(9): 686-689.]
- [5] Dong ZW, Qiao YL, Kong LZ, et al. The strategy and implementation for early detection of cancer in China [J]. China Cancer, 2008, 17(4): 256-263.[董志伟,乔友林,孔灵芝,等.中国癌症早诊早治的策略与实践[J]. 中国肿瘤, 2008, 17(4): 256-263.]
- [6] Song GH, Meng FS, Zhang XD. Collection and management of the data from population-based screening and treating program in cancer high incidence area [J]. China Cancer, 2010, 19(1): 36-38.[宋国慧,孟凡书,张向东.癌症高发发现场早诊早治项目点资料收集管理 [J]. 中国肿瘤, 2010, 19(1): 36-38.]
- [7] Lu JB. Commentary: cancer screening and early diagnosis and treatment of upper digestive tract in science and policy[J]. Advances in Clinical, 2014, 4(3): 35-41.[陆建邦.上消化道癌筛查及早诊早治的科学和策略[J]. 临床医学进展, 2014, 4(3): 35-41.]
- [8] Song GH. Application of Excel function for early diagnosis and early treatment of cancer screening database merge processing research [J]. Chinese Journal of Prevention and Control of Chronic Non-communicable Diseases, 2012, 20(2): 211-213.[宋国慧.应用 Excel 函数对癌症早诊早治筛查数据库合并处理的研究[J]. 中国慢性预防与控制, 2012, 20(2): 211-213.]
- [9] Wan XW, Wu Y, Shi YK. The problems and countermeasures in development of regional medical informatization in China[J]. Medicine & Philosophy, 2012, 33(13): 31-33.[万晓文,武媛,石应康.我国区域医疗信息化发展中存在的问题及对策[J]. 医学与哲学, 2012, 33(13): 31-33.]
- [10] Wang XD. Research on existing problems and countermeasures in current medical informatization[J]. Journal of Medical Informatics, 2011, 332: 44-47.[王晓丹.当前医疗信息化存在的问题及对策研究[J]. 医学信息学杂志, 2011, 332: 44-47.]
- [11] Chen WQ. Discussion on the clinical characteristics and trends of cancers in China according to cancer registry data [J]. Chinese Journal of Health Manage, 2016, 10(4): 249-251.[陈万青.从肿瘤登记数据看中国恶性肿瘤的发病特点和趋势[J]. 中华健康管理学杂志, 2016, 10(4): 249-251.]