

· 综述 ·

【编者按】 大数据作为重要的战略资源已经在全球范围达成共识,2015年10月,十八届五中全会提出实施国家大数据战略。如今,在城市建设、金融、电子商务等领域,大数据的应用随处可见,并改变着各行各业,而健康大数据在公共卫生领域中的应用却远远不足。今年国家卫生计生委网络安全和信息化工作领导小组全体会议提出积极推进健康医疗大数据应用示范,研究制定促进健康医疗大数据应用指导意见,推动健康医疗大数据依法有序安全开放,因此健康大数据在公共卫生领域有非常广阔的应用前景。本文重点阐述健康大数据在公共卫生领域中的应用与挑战,希望可以大力推进健康大数据在公共卫生领域中的应用。

健康大数据在公共卫生领域中的应用与挑战

孟润堂, 罗艺, 宇传华, 邱杰, 周达

【摘要】 随着卫生信息化建设进程的提速,医疗卫生与大数据正发生激烈“碰撞”,交叉形成了健康大数据这一新兴概念,健康大数据必将对未来民众健康产生深远影响。本文介绍了健康大数据的基本概念和特点,简要阐明了健康大数据的处理相关技术,重点阐述了健康大数据在疾病预测与预防、循证公共卫生决策、健康管理、健康监测与个性化医疗服务等方面的应用,指出了健康大数据当前所面临的诸如医学伦理学、关键技术突破、“误差”甄别等方面的挑战。以期为更好地推动大数据在公共卫生领域中的应用发挥积极指导作用。

【关键词】 数据挖掘; 公共卫生信息学; 健康大数据

【中图分类号】 R-056 【文献标识码】 A doi: 10.3969/j.issn.1007-9572.2015.35.029

孟润堂, 罗艺, 宇传华, 等. 健康大数据在公共卫生领域中的应用与挑战 [J]. 中国全科医学, 2015, 18 (35): 4388-4392. [www.chinagp.net]

Meng RT, Luo Y, Yu CH, et al. Application and challenges of healthy big data in the field of public health [J]. Chinese General Practice, 2015, 18 (35): 4388-4392.

Application and Challenges of Healthy Big Data in the Field of Public Health MENG Run-tang, LUO Yi, YU Chuan-hua, et al. School of Public Health, Wuhan University, Wuhan 430071, China

【Abstract】 With the speeding process of medical health informatization construction, "Fierce Collision" occurs between public health and big data, forming an emerging concept which is healthy big data. It will produce a profound influence on people's health in the future. This paper introduced the basic concepts and characteristics of healthy big data, briefly described the correlative technique of processing healthy big data in the area of public health, emphasized on the prediction and prevention of disease, evidence-based decision-making in public health, health management, health monitoring, personalized medical service and other aspects, and pointed out the main challenges, such as medical ethics, key technology breakthroughs, "errors" identification and so on, with a view to better promote big data application to play a positive guiding role in public health.

【Key words】 Data mining; Public health informatics; Healthy big data

大数据正在改变人们的生活及理解世界的方式,且更多的改变正蓄势待发^[1]。数据战略”上升为国家战略,认为大数据2012-03-22,奥巴马政府宣布投资2亿

美元拉动大数据相关产业发展,将“大数据”首次进入我国政府工作报告;2014-11-15,李克强总理提出在疾病防治、灾害预防、社会保障、电子政务等领域开展大数据应用示范。2015-08-19,国务院通过《关于促进大数据发展的行动纲要》(简称《纲要》),在全社会引起广泛影响。《科学》杂志于2014年底和2015年初分别刊登了“公共卫生遇上了大数据”和“将大数据纳入公共卫生系统^[2-3]”两篇文章,指出“强大的流行病学基础、稳健的知识整合、循证医学原

基金项目: 湖北省卫计委2014年“湖北省人口健康信息平台相关管理规范研究”子课题;湖北省卫计委“卫生计生统计工作规范研究”

作者单位: 430071 湖北省武汉市,武汉大学公共卫生学院(孟润堂,宇传华);宁波卫生职业技术学院护理学院(罗艺);武汉大学全球健康研究中心(宇传华);湖北省卫生计生信息中心(邱杰,周达)

通信作者: 宇传华, 430071 湖北省武汉市, 武汉大学公共卫生学院, 武汉大学全球健康研究中心; E-mail: yuchua@163.com

则以及拓展的转化研究议程”可以推动大数据在公共卫生方面的应用,这些均预示着公共卫生大数据研究的春天即将来临。大数据给人们带来的最直接利益就是对未来的预见,其可指导民众规避健康风险、预防疾病、提升生命质量。我国作为世界人口基数最大国,具有其他国家难以比拟的基础数据优势,海量公共卫生大数据亟待挖掘、整合、利用。

1 概念与特点

“大数据”是指成批大规模的、复杂的、可链接的数据信息。除基因组学信息外,还包括医疗信息、环境信息、金融信息、地理信息和社会媒体信息等^[2]。健康大数据是与维持机体健康或引起机体发生疾病/亚健康状态有关联的生活行为方式、遗传、社会环境因素及医疗过程中可以测量记录的数据信息,涉及某地区的所有人群。基于健康大数据可对全人群进行健康测量与评价,以形成正确的干预措施,促进人群健康。

大数据的特点可总结为海量(Volume)、多样(Variety)、快速(Velocity)、价值(Value)4个V^[4-5],即大、杂、快、疑等四个特点,具体解释见图1。

2 相关技术

大数据作为一种基础性资源,被处理后才能显现其潜在价值,那么如何更好地处理和利用大数据就显得非常重要。传统数据分析包括聚类分析、因子分析、相关分析、回归分析、数据挖掘算法等方法,通过挖掘、萃取、提炼以及整合数据,以最大化发挥数据的效用。

在分析大数据时传统数据分析方法依然适用^[6],但在处理较大规模的数据集时,效率跟不上用户预期,且处理非结构和半结构化复杂数据时面临一定困难。为了解决上述大数据问题,需采用非关系数据库(NoSQL)及云计算、云存储技术^[7-8]。如果数据是财富,那么大数据就是宝藏,而云计算就是挖掘和利用宝藏的利器。云计算的模式是业务模式,本质是数据的并行、分布式处理技术。数据是资产,云为数据资产提供存储、访问和计算。

在生物医学与健康研究中,Hadoop是可靠、高效、可伸缩的分布式处理软件框架,在目前大数据挖掘中最受欢迎且应

用广泛^[9]。Map Reduce则是一种可以用来并行处理大数据的编程模型,同一程序在Hadoop的框架下可以用各种不同语言(Java, Ruby, Python等)按Map Reduce的编程模型进行编写和运行。

大数据是一种新现象和新理念,是近年研究的一个技术热点。大数据与传统数据的区别见表1。大数据技术被设计用于在成本可承受的条件下,通过非常快速的搜集、存储、整理、分析,从海量、多样的数据中提取价值。其大数据平台技术架构见图2。

3 应用领域

目前,大数据在若干领域的运用中取得了较大成功,如天文学(斯隆数字巡天望远镜)、零售业(沃尔玛巨额交易量)、搜索引擎(Google基于已有网络数据的个人定制搜索),以及政治竞选(关

表1 大数据和传统数据特征比较

Table 1 Characteristics comparison between big data and traditional data

特征	大数据	传统数据
数据容量	不断增长中(TB、PB、ZB……)	MB、GB
处理时效	非常迅速(以s为单位)	较慢(以h或d为单位)
数据结构	半结构或非结构化	结构化
数据来源	完全分布式、云存储	中心式
数据整合	比较困难	相对容易
存储架构	Hadoop 分布式文件系统(HDFS),非关系数据库(NoSQL)	关系数据库管理系统(RDBMS)
接入方式	批处理或接近实时	交互式
分析对象	全体数据	样本数据
分析方法	描述分析为主	描述与推断相结合
分析结果	关联度、模式	可信区间、P值

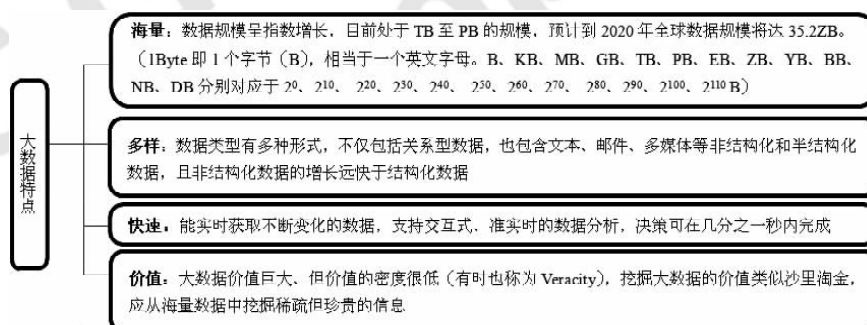
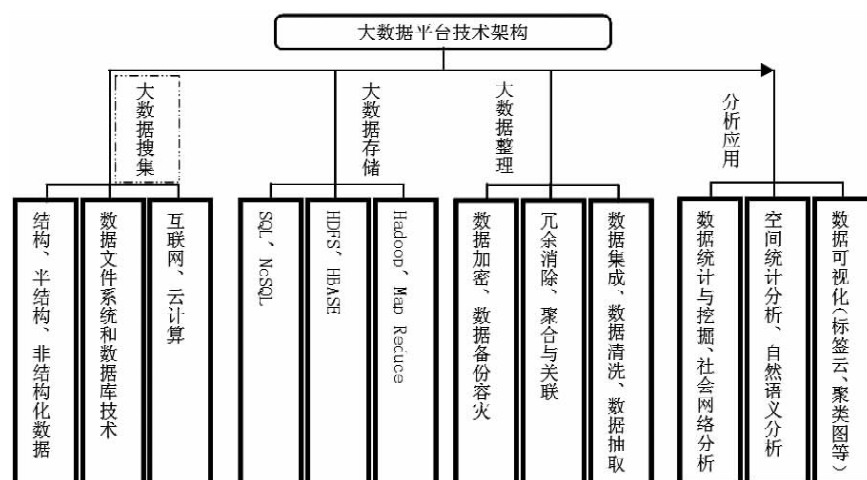


图1 大数据的4V特点

Figure 1 4V features of big data



注: 虚线框表示数据搜集过程可以简化省略, 如自由自愿贡献数据的众包(CrowdSourcing)搜集形式; SQL = 关系数据库, NoSQL = 非关系数据库, HDFS = Hadoop 分布式文件系统; HBASE = Hadoop Database, 是一个高可靠性、高性能、面向列、可伸缩的分布式存储系统

图2 大数据平台技术架构

Figure 2 Technology framework of big data

注竞选者最有可能基于网络搜索他们的支持候选人)等^[10],但在公共卫生领域的应用却远远不足,这使得健康大数据有非常广阔的应用前景。

3.1 疾病预测与预防 健康大数据可以帮助人们更好地预测、预防疾病。据估计,目前卫生保健工作者能确定的健康影响因素只有 10%~15%,剩下 85%~90% (包括健康行为、遗传、自然和社会经济环境因素等) 尚未知晓^[11]。一直以来预测未来疾病的发生极为困难,但利用大数据技术,基于处方药和非处方药的销售量、卫生服务咨询中心接到患者电话的数量和内容、关键词的点击量或搜索次数、社交网络浏览偏好等,使人群疾病预测成为可能。通过对全人群全程的信息跟踪,Google 曾于 2008 年底推出了一项大数据处理的“流感趋势”应用服务,该服务通过搜索引擎对关键词的监测,帮助人们了解美国境内不同地区的流感病例爆发情况^[12]。Gittelmann 等^[13]利用 Facebook 上的“喜好”数据,探讨了潜在健康结局的影响因素及其行为原因。通过主成分分析法和回归分析,控制年龄、种族、社会经济地位等变量后预测行为与健康状况的关系,显示“喜好”数据能提供更可靠、更及时和更具有成本-效益比的疾病预测结果,可作为传统公共卫生监测系统的补充。美国北卡罗莱纳州采用大数据技术开发的综合癌症信息与监测系统 (Integrated Cancer Information and Surveillance System, ICISS),将不同的数据、方法和系统有机结合起来,可从个人、健康服务商和地区卫生服务机构 (如该州癌症登记处、医疗保健、医疗救助、美国人口普查、疾控中心社区卫生指标和私人健康保险计划中心) 等不同途径,有效、及时、完整、准确地收集、整合和更新肿瘤相关资料。该系统不仅促进了人群肿瘤登记的发展,也为不同领域研究人员开展基于人群的肿瘤研究提供了可能和便利^[14]。通过人群健康大数据资料,可以实现疾病的预测与预防,减少重大疾病的发生和诊疗成本,全方位减缓疾病、促进健康。

3.2 循证公共卫生决策 浩如烟海的研究论文与研究报告蕴含着大量数据、证据、评论和概要,利用大数据技术与方法可为循证公共卫生决策提供准确有效的支

持。我国尚无独立的、专业的循证公共卫生决策数据库,目前主要依靠 Cochrane 协作网的循证决策数据库进行循证决策研究,并于 2005 年参与了全球第一个循证卫生决策网络——亚洲循证卫生决策网络 (EVIPNet - Asia) 的建设^[15]。近年来,循证医学的概念已经深入人心,其理论和方法已渗透到卫生决策和临床实践诸多方面,医疗卫生决策逐渐受到重视,但是国内循证公共卫生意识仍然很微弱,循证公共卫生思维难以形成,由于多种因素导致循证公共卫生决策的研究非常少。《柳叶刀》上有文章指出,目前在中国推动循证公共卫生政策遇到的最大障碍是研究者和政策制定者之间在许多认识上存在差异^[16]。系统评价的最大特征就是可重复性,是通过系统搜索和整合的方法产生结果,其中整合的过程是采用定量整合分析来进行。将个人数据集加入大数据能为循证医学提供最坚实的证据,能发现小样本无法发现的细微差别,为公共卫生决策者提供最新证据,指导卫生政策的制定或临床实践。例如,某研究者将饮用咖啡的生活习惯对前列腺癌的影响进行了剂量反应关系的 Meta 分析,结果表明:每天多饮 2 杯咖啡者患前列腺癌的风险降低了 2.5% [$RR = 0.975$, 95% $CI (0.957, 0.995)$]^[17]。此外,美国华盛顿大学健康测量与评价研究所 (IHME) 进行的全球疾病负担 (GBD) 研究项目,是利用大数据对世界人群进行全面健康测量与评价的最好案例,其数据库来源广泛 (包括 WHO、各国疾控中心、民政、统计、公安等部门,医疗机构以及专项调查等数据)^[18],可为政府合理分配卫生资源、形成正确公共卫生决策提供有价值的信息。但是,循证公共卫生决策领域的研究多为观察性研究或半试验性研究,从干预到结果出现的时间较长,且不可否认存在随机化不足或者不可行的问题^[19-20],在决策评价中通过引入工具变量、倾向评分匹配等方法,或者联系相关领域专家获取专业建议,将定性研究与定量研究相结合,可以使研究更为可信、可靠。国内建立完善的、专业的循证公共卫生决策数据库还有相当长的路要走,不能因为证据不多或干预取得效果的时间长而不去努力。利用大数据技术与方法无疑会加快循证公共卫生决策数据库建设的进程。

3.3 健康管理、健康监测与个性化医疗服务 利用健康大数据可以使研究者比以往任何时候都更好地对个体或人群进行健康管理、健康监测,并对不同个体提供差异化的医疗服务。健康管理是对个体或群体的健康进行全面的监测、分析、评估,提供健康咨询和指导以及对健康危险因素进行干预的全过程;健康档案承载着各种形式的健康大数据,并针对个体提供个体化健康管理服务^[21]。目前健康大数据的重要来源主要有两种,即电子健康档案 (Electronic Health Record, EHR) 和电子病历 (Electronic Medical Record, EMR)^[22-23],但不局限于以上两种形式^[21]。利用健康大数据技术与方法可将传统的健康数据 (如电子和纸质病历等) 与其他来源的个人数据 (如饮食、睡眠、锻炼习惯、生活方式、社交媒体和休闲、收入、教育等) 联系起来进行健康管理和监测^[21,24]。通过收集人体生理和行为的监测数据,积累构成含有健康状况和疾病风险重要信息的个体健康大数据,上传至云平台。这些数据包括智能的生理生化及行为传感器数据、求诊咨询用药数据、浏览和讨论数据、日常生活作息数据等。挖掘分析这些数据可以得到个人较为完整的健康状态及疾病预警信息 (尤其是针对个体在某一时期可能发生的重大疾病进行预警),结合个人基因谱和完整病史数据,将健康危险因素进行关联比对分析,跟踪病程进展、判断短期风险和长期预后,能够获得比临时求诊更准确的信息,从而进行更有效、更个性化的临床干预和健康指导。

健康监测是对个人健康进行全生命周期的管理,无论何时何地都可以访问相关信息,从而保证健康信息的完整性、连续性、实时性和预见性。例如微软开发的 HealthVault 网络平台健康云服务、苹果开发的 HealthKit 平台以及内置健康监测功能的 Apple Watch 等。人群健康信息以个体电子健康档案为载体转译进入公共卫生报告系统聚合成一定人口规模的健康大数据^[25]。

个性化医疗服务的最大特点是在个人实施健康管理基础上,通过对个人健康危险因素进行全面评估,制定具有差异化的健康促进计划。Map Reduce 和 Hadoop 分布式系统应用于临床大数据处理和分析,

给疾病诊断和个性化治疗开辟了新的途径,被认为是当前医学界的重大进展^[26]。如在肿瘤个性化治疗方面,美国临床肿瘤学会的“肿瘤学快速研究系统的多阶段计划(Cancer Lin Q)”使用可获得的开源和专有软件,对1万例乳腺癌患者的电子病历进行22项专项评估,内容包括10项肿瘤诊疗质量倡议(quality oncology practice initiative, QOPI)及其在治疗中应用的循证治疗方案。评估完成后,通过浏览和检索病历、产生假设、评价质量,为临床医生实时提供循证信息和治疗进展,并确定临床试验的参与资格,针对不同患者进行差异化治疗^[27]。但是,利用大数据不能固定个性化医疗服务,不能代替卫生保健专业人员来解释甚至简化诊疗程序,必须结合患者的教育程度、文化背景、社会支持系统及个人意愿来选择治疗方案^[28]。医务工作者只是提供个性化医疗服务项目,最终的选择仍然需要基于患者自身的具体情况。

4 面临的挑战与展望

Gartner发布的2014年新兴技术成熟度曲线显示,未来5~10年大数据技术将会成熟应用于各个领域^[29],健康大数据也将会快速应用到公共卫生领域之中,同时也面临一些挑战。

4.1 健康大数据使用中的安全、保密、共享、开放等医学伦理学问题 健康大数据不可避免地涉及人群的隐私信息,包括身体现状、健康史、个人信息,甚至基因、蛋白数据等,如若泄露,极可能会使患者个体的日常生活遭到难以预料的侵扰^[30]。个人隐私易被窃听的问题一直未能得到很好解决^[24]。若将数据加入到大数据库之前,通过电脑程序将能够被识别的患者个人信息从医疗记录中去除,理论上讲可以克服这个问题^[11,31]。但由于缺乏个体的识别信息,其他数据将无法和研究样本整合,难以证实因果行为和健康状况的关系,不能进行某类人群大范围的研究。通过特殊处理(如去识别化、数字身份加密等)可以较好地解决此问题,但仍绕不开信息识别,去识别化本身也需要处理可识别的信息,可能造成患者健康信息在不知情、未授权的情况下被他人盗用^[32];同样,数字身份加密在解密过程中也存在泄密盗用的可能。随着信息时代的深入,健康大数据更为专注于电子数据

的来源和信息技术手段,导致面临的伦理学问题挑战更为复杂多样^[33]。

此外,健康大数据的收集、存储、维护及使用方面,不仅涉及个人隐私问题,由于国家人口规模十分庞大、信息量巨大,还牵涉公众利益甚至国家安全。《纲要》中反复提及共享和开放的战略,强调由政府主导共享和开放数据,降低公众获取和利用政府数据资源的难度及成本,为公共卫生健康大数据研究铺平道路。与此同时,大数据意味着大责任、大伦理,任何单位或个人使用大规模健康大数据时均应该严格申请审查并备案,在法律允许的框架内使用相关数据,承担风险责任^[34]。健康大数据的使用过程既要破除壁垒,让信息互联互通;又要充分隐私保密,杜绝隐患。

4.2 突破大数据的关键技术,推动其在公共卫生中的应用 半结构化和非结构化数据量呈几何级数增长,传统的分析技术面临着较大的冲击和挑战。数据的广泛存在性使得数据越来越多地以不同的形式散布于不同的系统和平台之中^[35]。为了便于进行健康大数据分析,需要解决数据的多源异构性、数据的质量问题,各方面产生的大数据有待进行有效的整合。特别需要指出的是,在大数据时代虽然允许不精确的出现,但最基本、最重要的任务还是应该尽可能减少错误,保障质量。除上述技术挑战外,还有数据信息孤岛问题普遍存在,标准化难以实施等技术和非技术困难尚未得到有效彻底地解决。《纲要》指出,政府将推动公共数据互联互通、消除信息孤岛,避免重复建设和数据“打架”,整合各级平台。

4.3 甄别健康大数据使用中的“误差”,提高精度 大数据也会产生“大错误(Big Error)”^[2],流感在2013年最先袭击美国且造成十分严重的危害。当时科学家们先利用大数据技术,之后又采用传统的公共卫生监测方法分析流感的影响程度并进行估计,结果显示前者对流感的高峰期影响水平明显高估。“大数据”可以作为有效的工具来评估疾病负担和传播,Google流感趋势(Google Flu Trends, GFT)结合疾控中心的数据网络可以大幅提高预测性能,在流感传播和流行期间这一改进的模型可以更准确预测未来1周的感染情况^[36]。同样,卫生服务人员需要

认识到存在垃圾数据以及有责任维护数据的完整性和准确性^[28]。健康大数据使用者应认识到大数据本身不可能替代其他数据;虽然其弥补了很多以前数据的缺陷,但只是弥补性而不是取代性的功能,在疾病与健康预测方面甄别健康大数据的“误差”尤为重要。

一直以来,公共卫生与预防医学学科强调“三级预防”,但作为政府行为的“零级预防”非常关键。从星罗云布的大数据中通过分析快速获得有效信息,并及时为公共卫生决策提供依据,是一项基础的、长远的工作,需要政府层面的长期鼎力支持。健康大数据框架需要国家层次进行顶层设计,担负相应责任。建议政府加以科学引导,开放平台邀请更多社会资本参与,整合力量共建国家健康大数据研究中心。

通过数据的开放将医疗信息大数据直接传递给患者,从而改变目前将患者的记录存储在医疗机构,置患者于被动位置的模式。可以预见,在不久的将来,患者也会越来越多地参与到自己的健康管理之中,且在其中占主导地位。大数据无疑会对公共卫生领域有革命性的影响,通过大数据来识别健康影响因素,并采取相应干预措施促进人群健康,为个人或群体提供最适合的预防保健和治疗方式^[37];利用大数据能够促进新的发现,优化治疗效果,减少卫生支出。

为了实现健康大数据的巨大价值,公共卫生领域需要全面实现数据信息的标准化,增强数据的互用性,促进信息的共享,建立有效的数据管理方式,改进分析技术和方法,培养拥有公共卫生背景的数据分析专业人才。早日突破并解决上述挑战,大力推进健康大数据在公共卫生领域中的应用。

参考文献

- [1] Mayer - Schonberger V, Cukier K. Big data: a revolution that will transform how we live, work, and think [M]. Boston: Houghton Mifflin Harcourt, 2013.
- [2] Khoury MJ, Ioannidis JP. Medicine. Big data meets public health [J]. Science, 2014, 346 (6213): 1054 - 1055.
- [3] Fung IC, Tse ZT, Fu KW. Converting big data into public health [J]. Science, 2015, 347 (6222): 620.

- [4] Luna DR, Mayan JC, García MJ, et al. Challenges and potential solutions for big data implementations in developing countries [J]. Yearb Med Inform, 2014, 9 (1): 36-41.
- [5] Gantz J, Reinsel D. Extracting value from chaos [J]. IDC Iview, 2011 (1142): 9-10.
- [6] Chen M, Mao S, Liu Y. Big data: a survey [J]. Mobile Netw Appl, 2014, 19 (2): 171-209.
- [7] Merelli I, Pérez-Sánchez H, Gesing S, et al. Managing, analysing, and integrating big data in medical bioinformatics: open problems and future perspectives [J]. Biomed Res Int, 2014, 2014: 134023. doi: 10.1155/2014/134023.
- [8] Fan JH, Han F, Liu H. Challenges of big data analysis [J]. Natl Sci Rev, 2014, 1 (2): 293-314.
- [9] Peek N, Holmes J, Sun J. Technical challenges for big data in biomedicine and health: data sources, infrastructure, and analytics [J]. Yearb Med Inform, 2014, 9 (1): 42-47.
- [10] Murdoch TB, Detsky AS. The inevitable application of big data to health care [J]. JAMA, 2013, 309 (13): 1351-1352.
- [11] El-Emam K, Gray J, Greal M, et al. A policy forum on the use of big data in health care [EB/OL]. [2013-12-03]. <http://bipartisanpolicy.org/library/policy-forum-use-big-data-health-care/>.
- [12] Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data [J]. Nature, 2009, 457 (7232): 1012-1014.
- [13] Gittelman S, Lange V, Gotway Crawford CA, et al. A new source of data for public health surveillance: facebooklikes [J]. J Med Internet Res, 2015, 17 (4): e98.
- [14] Meyer AM, Olshan AF, Green L, et al. Big data for population-based cancer research [J]. NC Med J, 2014, 75 (4): 265-269.
- [15] WHO. EVIPNet Asia, Evidence-informed policy-making [EB/OL]. http://www.who.int/evidence/resources/country_reports/asia/en/.
- [16] Jiang F, Zhang J, Shen X. Towards evidence-based public health policy in China [J]. Lancet, 2013, 381 (9882): 1962-1964.
- [17] Liu H, Hu GH, Wang XC, et al. Coffee consumption and prostate cancer risk: a meta-analysis of cohort studies [J]. Nutr Cancer, 2015, 37 (3): 392-400.
- [18] 宇传华, 崔芳芳. 全球疾病负担研究及其对我国的启示 [J]. 公共卫生与预防医学, 2014, 25 (2): 1-5.
- [19] Brownson RC, Baker EA, Leet TL, et al. Evidence-based public health [M]. 2nd Revised ed. Oxford: Oxford University Press Inc, 2011: 10-11.
- [20] Tanaka S, Tanaka S, Kawakami K. Methodological issues in observational studies and non-randomized controlled trials in oncology in the era of big data [J]. Jpn J Clin Oncol, 2015, 45 (4): 323-327.
- [21] Kuziemyky CE, Monkman H, Petersen C, et al. Big Data in healthcare - defining the digital persona through user contexts from the micro to the macro: contribution of the IMIA Organizational and Social Issues WG [J]. Yearb Med Inform, 2014, 9 (1): 82-89.
- [22] Ross MK, Wei W, Ohno-Machado L. "Big Data" and the electronic health record [J]. Yearb Med Inform, 2014, 9 (1): 97.
- [23] Lissoyoy G. Big data meets the electronic medical record: a commentary on "identifying patients at increased risk for unplanned readmission" [J]. Med Care, 2013, 51 (9): 759-760.
- [24] Hill S, Merchant R, Ungar L. Lessons learned about public health from online crowd surveillance [J]. Big Data, 2013, 1 (3): 160-167.
- [25] Chute CG, Ullman-Muller M, Wood GM, et al. Some experiences and opportunities for big data in translational research [J]. Genet Med, 2013, 15 (10): 802-809.
- [26] Mohammed EA, Far BH, Naugler C. Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends [J]. Bio Data Mining, 2014, 7 (1): 22.
- [27] Sledge Jr GW, Miller RS, Hauser R. Cancer Lin Q and the future of cancer care [C] // American Society of Clinical Oncology educational book. American Society of Clinical Oncology, 2012: 430-434.
- [28] Ward JC. Oncology reimbursement in the era of personalized medicine and big data [J]. J Oncol Pract, 2014, 10 (2): 83-86.
- [29] Gartner. Gartner's 2014 hype cycle for emerging technologies maps the journey to digital business [EB/OL]. [2014-08-11]. <http://www.gartner.com/newsroom/id/2819918>.
- [30] Bellazzi R. Big data and biomedical informatics: a challenging opportunity [J]. Yearb Med Inform, 2014, 9 (1): 8-13.
- [31] Ward JC. Oncology reimbursement in the era of personalized medicine and big data [J]. J Oncol Pract, 2014, 10 (2): 83-86.
- [32] Rothstein MA. Is deidentification sufficient to protect health privacy in research? [J]. Am J Bioeth, 2010, 10 (9): 3-11.
- [33] Vayena E, Salathé M, Madoff LC, et al. Ethical challenges of big data in public health [J]. PLoS Comput Biol, 2015, 11 (2): e1003904.
- [34] Dove ES, Özdemir V. What role for law, human rights, and bioethics in an age of big data, consortia science, and consortia ethics? The Importance of Trustworthiness [J]. Laws, 2015, 4 (3): 515-540.
- [35] Khan N, Yaqoob I, Hashem IAT, et al. Big data: survey, technologies, opportunities, and challenges [J]. Scientific World J, 2014, 2014: 712826.
- [36] Davidson MW, Haim DA, Radin JM. Using networks to combine "Big Data" and traditional surveillance to improve influenza predictions [J]. Sci Rep, 2015 (5): 8154.
- [37] Mooney SJ, Westreich DJ, El-Sayed AM. Commentary: epidemiology in the era of big data [J]. Epidemiology, 2015, 26 (3): 390-394.

(收稿日期: 2015-06-20;

修回日期: 2015-10-20)

(本文编辑: 赵跃翠)