

国外慢性病患者风险预警研究现状*

王辅之^{1,2} 罗爱静² 付连国^{3△} 陈兴智¹ 潘 玮¹

(¹ 蚌埠医学院卫生管理系,安徽 蚌埠 233000;² 医学信息研究湖南省普通高等学校重点实验室(中南大学);湖南 长沙 410000;

³ 蚌埠医学院预防医学系,安徽 蚌埠 233000)

摘要 目的 慢性病患者风险预警问题是公共卫生领域的重要课题,也是卫生信息技术领域的研究热点。本文对国外慢性病患者风险预警研究现状进行梳理。采用文献回顾研究方法,检索 PubMed、EBSCO、EMBASE 3 个数据库收录的慢性病患者风险预警的相关研究论文,从“慢性病危险因素遴选”、“风险预警模型构建”和“慢性病风险预警模型质量评估”3 个方面,对慢性病患者风险预警建模所涉及的技术、方法和内容等问题进行归纳。慢性病的发生、发展与生物遗传、生活习惯、社会心理等因素密切相关;风险预警建模方法可采用基于统计学模型和人工智能的方法;预警模型的质量评估方法一般采用全局模型拟合评估、标度评估、受试者工作特征曲线(ROC)、净再分类改善指数(NRI)和整合辨识力改善指数(IDI)等。在今后的研究中,应该充分发挥公共卫生大数据优势,注重多因素联合作用对慢性病患者风险的影响,遵循循证医学的思想,并充分考虑慢性病患者风险预警中的伦理学问题。

关键词 慢性病;患病风险;预警模型

中图分类号:R554.1 **文献标识码:**A **文章编号:**1000-9760(2015)12-415-06

Review on the risk prediction model for chronic disease abroad

WANG Fuzhi, LUO Aijing, FU Lianguo, CHEN Xingzhi, PAN Wei

(Department of Health Management, Bengbu Medical College, Bengbu 233000, China)

Abstract: Objective Chronic disease risk prediction is an important issue in the fields of public health as well as health information technology. Present article reviews the current research situation of chronic disease risk prediction. Three database (PubMed\EBSCO\EMBASE) was retrieved for the literature collection of chronic disease risk prediction. The chronic disease risk prediction related technology, method, and content in three aspects included: chronic disease risk factors, risk prediction model, and quality assessment of prediction model was discussed. The occurrence and development of chronic diseases are closely related with biological genetic, living habits, and social psychological factors. Risk prediction model was based on two methods, the statistical model and the artificial intelligence model. The quality assessment of chronic disease risk prediction model commonly used global model fitness, calibration, receiver operating characteristic (ROC), net reclassification improvement (NRI), and integrated discrimination improvement (IDI). In future research, the public health big data should play more important role, and multi-factors combined risk prediction model should be best method for risk discrimination. The idea of evidence-based medicine and ethical issues should give full consideration in chronic risk prediction.

Keywords: Chronic disease; Illness risk; Prediction model

慢性非传染性疾病(non-communicable diseases, NCD)主要指心脑血管疾病、糖尿病、癌症和慢性阻塞性肺疾病为主的一系列疾病,已造成全球超

过 60% 的死亡和伤残,预计该数字到 2030 年将达到 75%^[1]。2011 年,第 66 届联合国大会预防和控制慢性病高级别会议通过了《关于预防和控制非传染性疾病的政治宣言》,该宣言是全球领导人首次对攻克心脏病、脑卒中、癌症、慢性呼吸系统疾病和糖尿病等慢性病所采取的具体行动达成共识^[2]。2013 年,第 66 届世界卫生大会又通过了《预防控

* [基金项目] 国家自然科学基金(青年科学基金)(编号:81502823);安徽省教育厅人文社会科学研究重点项目(编号:SK2015A404);安徽省高校优秀青年人才支持计划重点项目(编号:gxyqZD2016166)

△ [通信作者] 付连国, E-mail: 497366090@qq.com

制非传染性疾病预防全球行动计划(2013—2020)》^[3]。慢性病防控已成为全世界健康促进工作重点关注的內容。

慢性病预警是全球卫生保健领域新兴的研究课题,也是改善中低收入国家卫生保健状况的重要手段^[4]。近年来,国外慢性病患病风险预警问题的研究得到了众多科研机构和学者的广泛关注,而国内对于慢性病防控工作重点在“控”,而疏于“防”,慢性病患病风险预警研究尚未引起我国医疗卫生机构的足够重视^[5]。随着全世界卫生信息化进程的迅速发展和信息技术在卫生信息管理领域的快速普及,数字化管理产生了海量的医疗数据^[6]。基于公共卫生大数据系统,建立慢性病风险预警机制,对于降低慢性病患病风险、减少慢性病对人体的损伤,改善慢性病患者生命质量具有重要意义。

1 文献检索策略

本文以“chronic disease”、“risk”、“assessment”、“evaluation”、“prediction”为关键词对3个外文电子数据库(PubMed、EBSCO、EMBASE)中近十年来发表的相关文献进行检索,排除重复文献后,共检索文献116篇。为进一步准确获得所需文献,我们以常见的慢性病名称(“Coronary heart disease”、“Cardiovascular disease”、“Cancer”、“Diabetes”等)对检索结果的文献标题进行人工筛选,得到相关文献45篇,阅读摘要后排除不相关文献17篇,最终筛选出28篇关于各类慢性病患病风险预警的研究文献。文献纳入标准:文献涉及疾病类型属于国家卫计委等15个部门联合制定的《中国慢性病防治工作规划(2012—2015年)》中涉及的慢性病种类(包括心/脑血管疾病、糖尿病、癌症、慢阻肺、神经系统疾病、消化系统疾病等)。

通过阅读全文,对其中的文献按照慢性病危险因素遴选、风险预警模型构建和预警质量评估3部分进行整理,最后简述了国外相关的研究对我国今后开展慢性病患病风险预警研究的启示。

2 国外慢性病患病风险预警研究现状

2.1 慢性病危险因素遴选

慢性病患病风险指标的选择是慢性病患病风险预警的首要问题。慢性病的发生、发展与生物遗传、生活习惯、社会心理等因素密切相关。在日本,慢性病已更名为“生活习惯病”,也充分说明了慢性

病的罹患与病程发展不仅受到生理学、遗传学因素的影响,更与生活习惯、社会因素及心理因素密切相关。因此,慢性病风险指标应该综合考虑多因素对慢性病患病的综合影响作用^[7-8]。

评估危险因素与罹患慢性病之间的关系,必须保证患病风险模型中危险因素的可靠性和有效性。队列研究能够将患病时间因素纳入慢性病危险因素研究,因此,大多数慢性病危险因素的获取是基于队列研究得到的,但获得的数据一般以健康筛查或健康管理为目的,纳入的危险因素数量和慢性病患病评估结果一般是有限的,这又会影响最终慢性病患病预警模型的准确性^[9]。也有学者提出,在慢性病危险因素遴选中,应该充分发挥电子健康档案的作用,并考虑将健康危险因素数据的收集工作纳入个人电子健康档案架构中。慢性病风险预警模型变量的测量也应该设计标准化操作方法,而危险因素测量工具的效度和信度评估应该贯穿整个研究过程^[10]。

2.2 风险预警模型构建

慢性病风险预警模型的构建一般分为2种类型:基于统计学模型的风险预警和基于人工智能方法的风险预警。

在基于统计学模型的风险预警中,多元线性回归和多元logistic回归等回归模型是最常见的危险因素对疾病发生影响的分析方法。在给定了潜在的危险因素变量和慢性病患病情况后,统计模型中的回归系数就描述了危险因素与患病情况的关联强度,一旦回归系数确定,当前危险因素状况对个体慢性病患病风险的影响强度也随之确定^[11-12]。例如,评估个体心血管疾病患病风险预警模型“Framingham Risk Score”^[13]和乳腺癌患病风险预警模型“Breast Cancer Risk Assessment Tool(Gail model)”^[14]。

而基于人工智能方法的风险预警,则一般应用人工神经网络^[15]或支持向量机^[16]等工智能理论,对当前危险因素状况可能会给个体带来的危险因素给出评估得分。这种方法一般基于先验知识,将危险因素变量带入预警模型,由预警模型自动计算当前危险因素状况可能对个体健康带来的影响作用程度。

图1是一个典型的包含隐含层的BP神经网络拓扑模型,基于疾病风险因素与健康状况的先验知识,对该网络进行训练,并设定最大误差允许值,当训练误差满足最大允许值后,即认为网络训练结

束。将疾病风险因素作为网络输入,该网络会模拟专家判断,给出患病风险最佳拟合结果并输出,完成疾病风险预警评估。

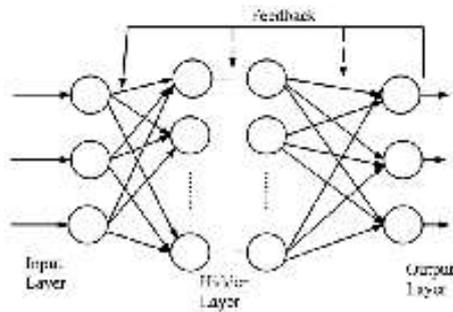


图 1 基于 BP 神经网络的疾病风险预警模型^[17]

2.3 慢性病风险预警模型的质量评估

预警质量评估也是开展慢性病患病风险预警模型研究的重要课题。一个好的风险预警模型能够根据输入变量情况,对慢性病患病风险给出准确预测。常用的疾病风险预警质量评估方法包括:全局模型拟合评估、标度评估、受试者工作特征曲线(ROC)等。

2.3.1 全局模型拟合 全局模型拟合(global model fitness)评估的目的是基于最少变量发现最佳拟合模型。赤池信息量准则(Akaike information criterion, AIC)和贝叶斯信息准则(Bayesian information criterion, BIC)是全局模型拟合评估常用的 2 种方法^[18]。

AIC 准则是衡量统计模型拟合优良性的一种标准,用于权衡估计模型的复杂度和模型拟合数据的优良性^[19]。一般表示为:

$$AIC = (2k - 2L) / n$$

其中 k 是参数的数量, L 是对数似然值, n 是观测值数量。k 小意味着模型简洁, L 大意味着模型精确,因此,该评估模型兼顾了评估的简洁性和精确性。

而贝叶斯信息准则^[20]则是在信息缺失情况下,对部分未知状态用主观概率估计,然后用贝叶斯公式对发生概率进行修正,最后利用期望值和修正概率给出最优决策的评价方法。

2.3.2 标度 标度(calibration)是对不同分组样本,疾病风险预测结果与观测值相比较得到的差值,可以通过图形直观观测。标度越小,表明疾病风险预测结果与观测值越接近,其统计学显著性通常采用 Hosmer-Lemeshow's 卡方检验来评价^[21]。图 2 是 2 种急性生理功能评分模型(simplified a-

cute physiology score II, SAPS-II)对 ICU 病人死亡情况的预测评价结果的标度散点图^[20],从图中可以看出相比传统 SAPS-II 预测模型, NIR-SAPS 预测模型的预测风险更低。

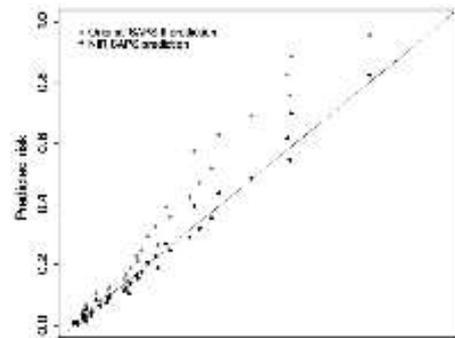
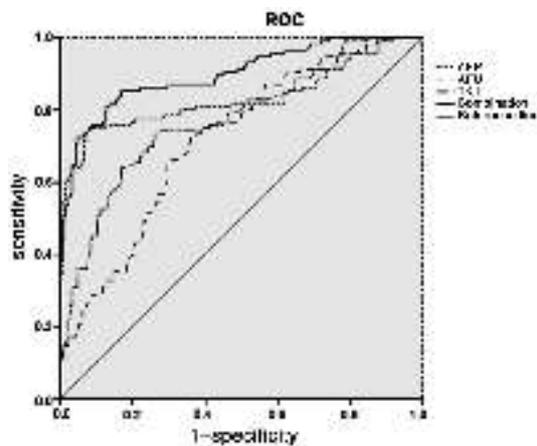


图 2 2 种 SAPSII 对 ICU 病人死亡情况的预测^[22]

2.3.3 受试者工作曲线 受试者工作曲线(receiver operating characteristic curve, ROC)是根据一系列不同的二分类变量,以真阳性率(灵敏度)为纵坐标,假阳性率(1-特异度)为横坐标绘制的曲线,对同一种疾病的两种或两种以上诊断方法进行比较。可将各试验的 ROC 曲线绘制到同一坐标中,直观地鉴别优劣。靠近左上角的 ROC 曲线所代表的受试者工作最准确,也可通过分别计算各个试验的 ROC 曲线下的面积(AUC)进行比较, AUC 最大的评价模型,其试验诊断价值最佳。ROC 曲线分析是疾病风险预警模型效果评价的重要工具^[23]。图 3 是分析 AFU、AFP、TK1 和 3 种标志物联合模型对肝癌细胞诊断效能的 ROC 曲线。从图中也可明显看出,3 种标志物联合模型的肝癌细胞诊断效能最高^[24]。



AFU、AFP、TK1 和联合模型的 AUC 分别为: 0.781, 0.832, 0.773 和 0.900^[24]

图 3 4 种模型对肝癌细胞诊断效能的 ROC 曲线

2.3.4 其他疾病风险预警模型质量评估方法 还有一些其他的疾病风险预警模型质量评估方法,如净再分类改善(net reclassification improvement, NRI)指数^[25]和整合辨识力改善(integrated discrimination improvement, IDI)指数^[26],在疾病风险预警模型质量评估中也被广泛应用。相对于没有纳入CGS的风险预警模型,加入CGS的预警模型对2型糖尿病组和健康对照组的识别率分别提

高4.36%和6.99%,总识别率提高11.4%^[27]。表1是增加8个单核苷酸多态性(SNPs)的联合基因评分(combined genetic score, CGS)预警模型与没有增加CGS的预警模型对中国人群二型糖尿病患病风险的预测评估结果的NRI指数计算结果。使用8个SNP联合基因预警模型对于预测2型糖尿病的辨识能力超越了临床危险因素(性别、年龄和BMI)对2型糖尿病风险的辨识能力^[28]。

表1 基于NRI的2型糖尿病患病风险预警效果比较

Predicted risk(without CGS)	Reclassified predicted risk(with CGS)					% (N) of subjects reclassified with		Net correctly reclassified(%)
	<5%	5%~	10%~	15%~	≥20%	increased risk	decreased risk	
T2D patients(N=5820)								
<5%	252	93	9	0	0	22.2%	17.8%	4.36%
5%~	139	678	250	87	17	(1290)	(1036)	
10%~	0	307	460	292	169			
15%~	0	25	261	283	373			
≥20%	0	0	69	235	1821			
Healthy controls (N=2560)								
<5%	1151	41	0	0	0	10.1%	17.1%	6.99%
5%~	114	207	64	8	0	(258)	(437)	
10%~	1	128	122	66	14			
15%~	0	19	73	72	65			
≥20%	0	0	34	68	313			
Net reclassification improvement(95% CI)						11.4(7.7~15.1)		
						P<0.001		

3 慢性病风险预警模型未来的研究方向

3.1 充分发挥公共卫生大数据优势

尽管对于大数据的定义普遍强调其数量的“大”,但国外有研究认为大数据具备“4V”属性:大量(volume)、多样(variety)、高速(velocity)和真实(veracity)^[28-29]。卫生大数据正在公众健康促进工作中发挥越来越重要的作用^[30],基于大数据的慢性病相关研究也得到广泛关注。在加拿大,Tu等学者通过链接多个数据库,使得心血管门诊研究数据形成一个强大的数据库,从而为改善病人的心血管健康状况提供支持^[31]。Shaikh等学者也探讨了如何融合社会技术框架(sociotechnical frameworks)、信息平台和健康相关政策,从而促进数据流动与创新,建立基于流行病学理论的癌症预防与

控制方法^[32]。目前,国内对医疗大数据利用仍处于起步阶段,虽然建立了国家层面的传染病预警机制,但在基于公共卫生大数据开展慢性病预警方面,如何利用公共卫生大数据,从中挖掘未知的慢性病危险因素,构建预警模型,开展大范围的、实时的慢性病患者风险预警工作,需要今后进一步完善。

3.2 注重多因素联合作用对慢性病患者风险的影响

传统的慢性病预警由于受到研究者知识结构和关注角度的不同,多从单纯的社会心理、生活行为习惯或生物基因、遗传等渠道寻找可能的慢性病发病因素。然而大量研究表明采用多维度因素联合评估能够更好地完成慢性病风险预警^[33]。本文2.3.4节提及的2型糖尿病风险预警模型就是结

合 8 个 SNP 联合基因及社会人口学因素共同构建预警模型^[27],取得较好的辨识结果。中国疾病预防控制中心慢性非传染性疾病预防控制中心创造性地引进世界卫生组织阶梯式检测(STEPWise)方法,建立了中国慢性病及其危险因素检测系统。该系统采用末端抽样方法,采用问卷调查、身体测量和生化指标检测,很好地覆盖了 WHO 阶梯式检测 3 阶段的全部内容,为我国成年居民常见慢性病的流行情况与变化趋势提供了大量基础数据^[34]。这种多因素联合作用,完成对疾病风险预警的研究思路,应该在未来的慢性病患者风险预警研究中得到更广泛的应用。

3.3 遵循循证医学的思想

无论采取何种风险预警模型,科学、有效的慢性病危险因素筛选是风险预警模型成功的关键。目前对慢性病的危险因素,学术界基本形成明确共识。但随着大数据应用的拓展和医学遗传学的进步,以及人们对慢性病发病机制、致病机理不断产生新的认识,这种新认识不仅存在于生物医学领域,也存在于社会心理学领域^[35],但都应遵循循证医学思想,基于流行病学调查数据的统计分析结果和临床研究证据(包括大样本随机对照临床试验、系统性评价和 Meta 分析),完成慢性病危险因素的筛选。

3.4 充分考虑慢性病患者风险预警中的伦理学问题

公共卫生领域的伦理学问题由来已久,早在 20 世纪早期,Biggs 教授就提出“尊重个人隐私、解决信息保密问题”是公共卫生报告可行化的唯一途径^[36]。大数据时代,公共卫生领域面临的伦理问题更加突出。有学者提出了数字化疾病侦测(digital disease detect,DDD)的概念^[37],认为 DDD 与传统的流行病学目标相同,但仅专注于电子化的数字卫生资源。瑞士苏黎世大学的 Effy Vayena 教授提出 DDD 所面临的伦理学挑战有 3 个维度:“内容敏感性”、“道德规范和方法的联系”和“合法性要求”,具体包括:数据的商业应用、用户许可、流行病学参与、数据起源等 9 个方面的挑战^[38]。这些相关研究成果对我国今后开展慢性病风险预警研究具有重要的参考价值。

4 结语

慢性病患者风险预警研究是我国公共卫生领

域的重要课题,也是我国医学信息数字化进程的重要应用。随着大数据时代的到来和人民群众健康观念的转变,慢性病患者风险预警问题将在更大范围内引起公众的关注。本文初步介绍了国外慢性病患者风险预警研究现状及所涉及的技术、方法和内容等问题,希望为我国今后相关研究的顺利开展提供借鉴。

参考文献:

- [1] WHO. World health report 2002[R]. Geneva:WHO,2002.
- [2] United Nations General Assembly. Political declaration of the high-level meeting of the general assembly on the prevention and control of non-communicable disease[Z]. New York, 2011.
- [3] WHO. Updated revised draft global action plan for the prevention and control of non-communicable diseases 2013-2020 [Z]. Geneva,2013.
- [4] Wyber R, Vaillancourt S, Perry W, et al. Big data in global health:improving health in low-and middle-income countries [J]. Bull World Health Organ,2015,93(3):203-208.
- [5] 马家奇. 中国疾病预防控制中心信息体系规划与发展[J]. 中国数字医学,2011,6(6):11-13.
- [6] 马家奇. 公共卫生大数据应用[J]. 中国卫生信息管理,2014,2(11):174-181.
- [7] Moons K G, Kengne A P, Woodward M, et al. Risk prediction models: development, internal validation, and assessing the incremental value of a new (bio) marker[J]. Heart,2012,98(11):683-690.
- [8] Toll D B, Janssen K J, Vergouwe Y, et al. Validation, updating and impact of clinical prediction rules:a review[J]. J Clin Epidemiol,2008,61(2):1085-1094.
- [9] Kim H C. Clinical utility of novel biomarkers in the prediction of coronary heart disease[J]. Korean Circ J, 2012, 42: 223-228.
- [10] Oh S M, Stefani K M, Kim H C. Development and Application of Chronic Disease Risk Prediction Models[J]. Yonsei Med J, 2014, 55(4): 853-860.
- [11] Aaron S D, Stephenson A L, Cameron D W, et al. A statistical model to predict one-year risk of death in patients with cystic fibrosis[J]. J Clin Epidemiol, 2015, 68(11): 1336-1345.
- [12] Vargas A C, Da Silva L, Lakhani S R. The contribution of breast cancer pathology to statistical models to predict mutation risk in BRCA carriers[J]. Fam Cancer, 2010, 9(4): 545-553.
- [13] Reddy R K, Mahendra J, Gurumurthy P, et al. Identification of predictable biomarkers in conjunction to framingham risk score to predict the risk for cardiovascular disease (CVD) in non cardiac subjects[J]. J Clin Diagn Res, 2015, 9(2): 23-27.

- [14] Tehranifar P, Protacio A, Akinyemiju T F, et al. Acculturation and ethnic variations in breast cancer risk factors, gail model risk estimates and mammographic breast density[J]. *Cancer Epidemiol Biomarkers Prev*, 2015, 24(4):760.
- [15] Chen Y C, Ke W C, Chiu H W. Risk classification of cancer survival using ANN with gene expression data from multiple laboratories[J]. *Comput Biol Med*, 2014, 48(3):1-7.
- [16] Sen I, Saraclar M, Kahya Y P. Computerized diagnosis of respiratory disorders. SVM based classification of VAR model parameters of respiratory sounds [J]. *Methods Inf Med*, 2014, 53(4):291-295.
- [17] Xie N N, Hu L, Ti TH. Lung cancer risk prediction method based on feature selection and artificial neural network[J]. *Asian Pac J Cancer Prev*, 2014, 15(23):10539-10542.
- [18] Penny W D. Comparing dynamic causal models using AIC, BIC and free energy[J]. *Neuroimage*, 2012, 59(1):319-330.
- [19] McGeechan K, Macaskill P, Irwig L, et al. Assessing new biomarkers and predictive models for use in clinical practice: a clinician's guide[J]. *Arch Intern Med*, 2008, 168(21):2304-2310.
- [20] Harrell F E. Regression modeling strategies; with applications to linear models, logistic regression, and survival analysis [M]. New York:Springer, 2001.
- [21] Paul P, Pennell M L, Lemeshow S. Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets [J]. *Stat Med*, 2013, 15(32):67-80.
- [22] Haaland O A, Lindemark F, Flaatten H, et al. A calibration study of SAPS II with Norwegian intensive care registry data [J]. *Acta Anaesthesiol Scand*, 2014, 58(6):701-708.
- [23] Mirmiran P, Esmailzadeh A, Azizi F. Detection of cardiovascular risk factors by anthropometric measures in Tehranian adults: receiver operating characteristic (ROC) curve analysis [J]. *Eur J Clin Nutr*, 2004, 58(8):1110-1118.
- [24] Zhang S Y, Lin B D, Li B R. Evaluation of the diagnostic value of alpha-L-fucosidase, alpha-fetoprotein and thymidine kinase 1 with ROC and logistic regression for hepatocellular carcinoma[J]. *FEBS Open Bio*, 2015, 5:240-244.
- [25] Leening M J, Vedder M M, Witteman J C, et al. Net reclassification improvement: computation, interpretation, and controversies; a literature review and clinician's guide[J]. *Ann Intern Med*, 2014, 160(2):122-131.
- [26] Kerr K F, McClelland R L, Brown E R, et al. Evaluating the incremental value of new biomarkers with integrated discrimination improvement[J]. *Am J Epidemiol*, 2011, 174(3):364-374.
- [27] Jee K, Kim G H. 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.
- [28] Tam C H, Ho J S, Wang Y, et al. Use of net reclassification Improvement (NRI) method confirms the utility of combined genetic risk score to predict type 2 diabetes[J]. *PLoS One*, 2013, 8(12):e83093.
- [29] Miller A R, Tucker C. Health information exchange, system size and information silos[J]. *J Health Econ*, 2014, 33(1):28-42.
- [30] Otero P, Hersh W, Jai Ganesh A U. Big data; are biomedical health informatics training programs ready[J]. *Yearb Med Inform*, 2014, 9(1):177-181.
- [31] Tu J V, Chu A, Donovan L R, et al. The Cardiovascular health in ambulatory care research team (CANHEART); using big data to measure and improve cardiovascular health and healthcare services[J]. *Circ Cardiovasc Qual Outcomes*, 2015, 8(2):204-212.
- [32] Shaikh A R, Butte A J, Schully S D, et al. Collaborative biomedicine in the age of big data; the case of cancer[J]. *J Med Internet Res*, 2014, 16(4):e101.
- [33] Lebrón-Aldea D, Dhurandhar E J, Pérez-Rodríguez P, et al. Integrated genomic and BMI analysis for type 2 diabetes risk assessment[J]. *Front Genet*, 2015, 75(6):1-29.
- [34] 元晓, 张普洪, 张勇, 等. 中国慢性病预防控制策略现状[J]. *中国慢性病预防与控制*, 2012, 2(20):214-217.
- [35] Stanton A L, Revenson T A, Tennen H. Health psychology: psychological adjustment to chronic disease [J]. *Annu Rev Psychol*, 2007, 58(1):565-592.
- [36] Winslow C E. The life of Hermann M Biggs: physician and statesman of the public health [M]. Philadelphia, PA: Lea and Fe bigger, 1929:120.
- [37] Brownstein J S, Freifeld C C, Madoff L C. Digital disease detection--harnessing the Web for public health surveillance [J]. *N Engl J Med*, 2009, 360(21):2153-2155, 2157.
- [38] Effy Vayena, Marcel Salathe, Lawrence C, et al. Ethical challenges of big data public health[J]. *PLoS Comput Biol*, 2015, 11(2):e1003904.

(收稿日期 2015-11-04)