

## • 急救大数据 •

## 急救数据库建设初探

赵宇卓 王俊梅 潘菲 李沛尧 贾立静 李开源 冯聪

刘同波 张政波 曹德森 黎檀实

100853 北京,解放军总医院急诊科(赵宇卓、潘菲、贾立静、李开源、冯聪、黎檀实),医学工程保障中心(李沛尧、张政波、曹德森),计算机应用与管理科(刘同波),医疗大数据中心(张政波);  
100191 北京航空航天大学大学生物与医学工程学院(王俊梅)

通讯作者:黎檀实,Email: lts301@163.com

DOI: 10.3760/cma.j.issn.2095-4352.2018.06.022

**【摘要】 目的** 建立急救患者多病种数据库,为研究急救领域临床科研问题提供基于“真实世界”的数据信息。**方法** 从解放军总医院的医院信息系统(HIS)、实验室信息系统(LIS)、急诊专科系统、急诊护理系统及床旁监护设备内,提取 2014 年 1 月至 2018 年 1 月在急诊科就诊患者的结构化和非结构化信息,通过建立表单、编写代码建立数据库,并对数据进行处理。**结果** 解放军总医院初步建立了单中心数据库,即急救数据库。共收集 2014 年 1 月至 2018 年 1 月在解放军总医院急诊科就诊的 530 585 例次患者的分诊信息以及 22 941 例次急诊抢救单元收治患者的信息,包括人口统计学资料、预检分诊情况、诊疗记录、生命体征、实验室检验结果、影像学检查报告、胃肠镜检查报告等信息;建立了 12 个包含患者不同类别信息的表单,包括 PATIENTS, TRIAGE\_PATIENTS, EMG\_PATIENTS\_VISIT, VITAL\_SIGNS, CHARTEVENTS, MEDICAL\_ORDER, MEDICAL\_RECORD, NURSING\_RECORD, LAB\_TEST\_MASTER, LAB\_RESULT, MEDICAL\_EXAMINATION, EMG\_INOUT\_RECORD, 包含了患者在急诊就诊过程中绝大部分临床信息。**结论** 高质量急救数据库的建立能为基于数据技术的科学研究提供坚实基础,急救数据库的建立模式可为其他医疗机构多病种数据库的建设提供参考。

**【关键词】** 急救; 数据库; 信息化; 数据挖掘

**基金项目:** 国家自然科学基金(81701961); 军事医学创新专项(14CXZ005); 军队后勤科研计划项目(AWS15J004); 北京市科技新星计划项目(XX2018019); 解放军总医院医疗大数据科研项目(2017MBD-30); 医疗大数据应用技术国家工程实验室(2017-148)

**Pilot research: construction of emergency rescue database** Zhao Yuzhuo, Wang Junmei, Pan Fei, Li Peiyao, Jia Lijing, Li Kaiyuan, Feng Cong, Liu Tongbo, Zhang Zhengbo, Cao Desen, Li Tanshi  
Department of Emergency, Chinese PLA General Hospital, Beijing 100853, China (Zhao YZ, Pan F, Jia LJ, Li KY, Feng C, Li TS); Department of Engineering and Maintenance Center, Chinese PLA General Hospital, Beijing 100853, China (Li PY, Zhang ZB, Cao DS); Department of Computer Application and Management, Chinese PLA General Hospital, Beijing 100853, China (Liu TB); Medical Information Center, Chinese PLA General Hospital, Beijing 100853, China (Zhang ZB); School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China (Wang JM)  
Corresponding author: Li Tanshi, Email: lts301@163.com

**【Abstract】 Objective** To construct a database containing multiple kinds of diseases that can provide "real world" data for first-aid clinical research. **Methods** Structured or non-structured information from hospital information system, laboratory information system, emergency medical system, emergency nursing system and bedside monitoring instruments of patients who visited department of emergency in PLA General Hospital from January 2014 to January 2018 were extracted. Database was created by forms, code writing, and data process. **Results** Emergency Rescue Database is a single center database established by PLA General Hospital. The information was collected from the patients who had visited the emergency department in PLA General Hospital since January 2014 to January 2018. The database included 530 585 patients' information of triage and 22 941 patients' information of treatment in critical rescue room, including information related to human demography, triage, medical records, vital signs, lab tests, image and biological examinations and so on. There were 12 tables (PATIENTS, TRIAGE\_PATIENTS, EMG\_PATIENTS\_VISIT, VITAL\_SIGNS, CHARTEVENTS, MEDICAL\_ORDER, MEDICAL\_RECORD, NURSING\_RECORD, LAB\_TEST\_MASTER, LAB\_RESULT, MEDICAL\_EXAMINATION, EMG\_INOUT\_RECORD) that containing different kinds of patients' information. **Conclusions** The setup of high quality emergency databases lay solid ground for scientific researches based on data. The model of constructing Emergency Rescue Database could be the reference for other medical institutions to build multiple-diseases databases.

**【Key words】** Emergency rescue; Database; Information; Data mining

**Fund program:** National Natural Science Foundation of China (81701961); Military Medical Innovation Project (14CXZ005); Military Logistics Science Study Planning Program (AWS15J004); Science and Technology New Star Program of Beijing (XX2018019); Big-data Research and Development Project of Chinese PLA General Hospital (2017MBD-30); National Engineering Laboratory for Industrial Big-data Application Technology (2017-148)

近20年来,随着信息化进程的普及和加快,大多数医疗机构不仅逐步实现了从纸质医疗记录向电子医疗记录的跃迁<sup>[1]</sup>,且数据总量呈“爆发”式增加。数据的类型多种多样,除医疗机构收集的数据外,还有基因组学数据、公共卫生领域大数据、可穿戴设备采集的生命体征数据及开放的互联网数据,且均呈现迅速增长趋势。传统的数据保存方法已不能满足当前“数据爆发”时代的需求,随着计算机技术、信息化技术的不断发展,以数据库作为载体的数据储存模式应运而生,成功打破了数据存储的桎梏<sup>[2]</sup>。但真正要想使数据库中庞大体量的数据发挥作用,就必须应用先进的科技手段。随着医学的发展,新型疾病评分方法层出不穷,医疗界越来越重视应用现代技术实现对疾病的准确评估和预警,例如人工智能心电分析系统<sup>[3]</sup>。而基于数据科学技术,使既往诸多复杂且耗时的临床科学研究都可在短期内得到解决<sup>[4-5]</sup>,例如:Joudaki等<sup>[6]</sup>通过数据挖掘技术监测医保欺诈行为;Zimmerman等<sup>[7]</sup>通过患者术前心肌损伤标志物预测术后病死率;段立伟等<sup>[8]</sup>基于Logistic回归模型构建联合预测因子预测脓毒症患者预后。医疗大数据库建设及数据分析技术正在深刻地改变着传统的医学研究、临床医疗、卫生管理支持模式,并已成为大样本临床研究、个性化诊疗、精准医疗、疾病监测预警、卫生经济评价、政府政策评估、新药研制等医学科技新进展的重要支撑手段。为此,本研究在原有信息化建设基础上,旨在完成急救数据库建设,为急救领域相关研究提供“真实世界”的数据信息,进而推动急救医学发展。

## 1 资料与方法

**1.1 数据来源:**数据源自2014年1月至2018年1月解放军总医院急诊科就诊患者的真实信息,产生及存储以上信息的途径包括解放军总医院医院信息系统(HIS)、实验室信息系统(LIS)、急诊专科系统、急诊护理系统以及床旁监护设备。

**1.2 数据分类:**根据临床科研需要,拟采集的数据包括患者人口统计学资料、预检分诊情况、诊疗记录、生命体征、实验室检验及其他辅助检查结果报告。

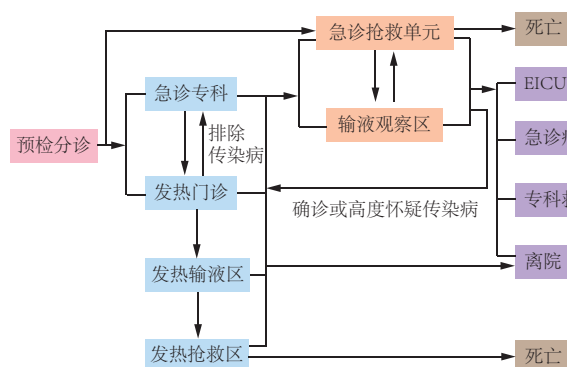
**1.3 数据先期处理:**将患者基本信息及诊疗信息录入数据库前进行去隐私化处理。主要包括:①隐藏患者真实姓名,并将患者真实ID号(患者在解放军总医院就诊期间所产生的唯一身份ID)进行复杂加密处理,避免此类可直接指向特定患者的信息出现在数据库中,其他信息均为患者诊疗过程中产生的真实信息;②隐藏患者及家属的联系电话、详细家庭地址等信息,避免此类敏感信息出现在数据库中。

**1.4 数据库建立:**在Oracle数据库建立过程中,参考并借鉴国外大型多参数数据库建设经验,如重症监护医学信息数据库Ⅲ(MIMIC-Ⅲ)等<sup>[9]</sup>。设置数据库基本信息,根据患者信息类别建立包含患者参数信息的多个表单,建立和确定表单主键和外键,使不同表单信息互相关联。对数据库中信息进行规范化处理,并删除部分无用数据避免冗余,达到数据清洗的目的。基于已建立的数据库提取患者部分信息,并进行可视化分析。

**1.5 统计学处理:**使用Excel 2016对数据进行统计分析以及可视化展示;利用VLOOKUP函数通过患者ID号进行表单之间的关联和匹配;利用数据透视表,从不同角度对数据进行分类汇总。

## 2 结果

**2.1 急诊患者就诊流程(图1):**为便于了解数据库中数据来源及框架,首先介绍患者于解放军总医院急诊科就诊基本流程。患者首先经预检分诊,被划分到不同专科,包括急诊综合、急诊腹痛、急诊创伤、急诊神经内科、急诊儿科、急诊眼科、急诊耳鼻喉科、急诊骨科、发热门诊(包括肠道门诊)等科室,危重患者经分诊直接送入抢救单元救治;根据进一步诊疗,急诊专科患者可转入输液观察区或抢救单元;抢救单元患者经治疗平稳后可转入输液观察区。发热患者转入发热门诊就诊,部分患者经排除传染病后可转入常规急诊系统救治;急诊系统患者确诊或高度怀疑传染病会转到发热门诊。输液观察区及抢救单元患者去向为急诊住院、专科救治或离院。



注: EICU 为急诊重症加强治疗病房

图1 解放军总医院急诊患者就诊流程

**2.2 急救数据库患者特点:**急救数据库中包含2014年1月至2018年1月解放军总医院急诊科530585例次患者的就诊分诊信息以及22941例次急诊抢救单元收治的患者信息;所有患者年龄0~117岁(图2)。抢救单元收治患者比例占全部就诊患者的4.32%,其中死亡患者540例(2.35%);死亡患者年龄分布见图3;死亡患者住院时间中位数为4h,存活患者为12h。图4显示,急诊各专科患者就诊比例最多为急诊儿科(19.14%),其次为急诊综合科(17.11%)、急诊创伤科(6.99%)。

**2.3 急救数据库数据分类(表1):**经数据处理,目前急救数据库所包含数据分类为:①患者人口统计学信息:包括性别、出生日期等;②预检分诊信息:包括分诊时间、分诊级别、分诊区域、分诊科室、来诊方式等;③诊疗信息:包括诊断信息(根据国际疾病分类ICD-9进行规范化处理)、医疗文书、护理文书、医嘱信息等;④生命体征信息:包括患者的心率、呼吸频率、体温、收缩压、舒张压、血氧饱和度等(部分信息经护理人员确认后记录,部分为床旁监护仪自动采集);⑤实验室检验信息:包括患者在救治阶段所采集的

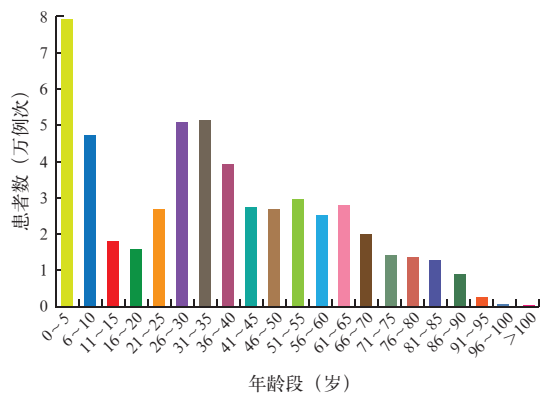


图2 解放军总医院急救数据库中 2014 年 1 月至 2018 年 1 月 530 585 例急诊就诊患者的年龄分布

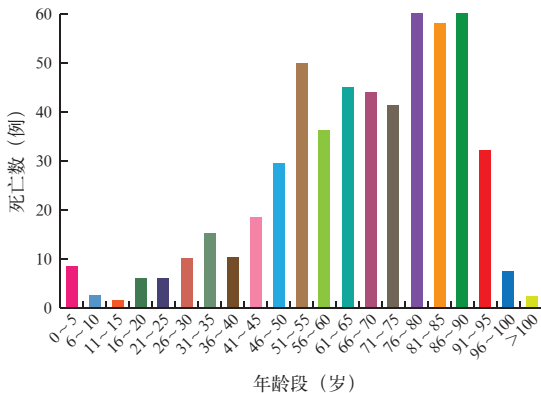


图3 解放军总医院急救数据库中 2014 年 1 月至 2018 年 1 月急诊抢救单元收治 540 例死亡患者的年龄分布

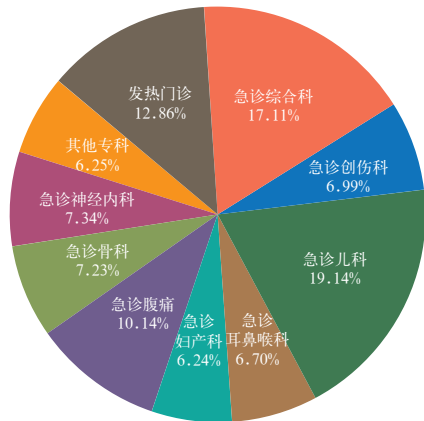


图4 解放军总医院急救数据库中 2014 年 1 月至 2018 年 1 月急诊各专科患者就诊比例

血、尿、粪、痰液、脑脊液、分泌物等标本的化验结果；⑥ 其他辅助检查信息：包括患者救治期间所产生的影像学、胃镜、肠镜、支气管镜、病理等辅助检查结果报告。

**2.4 急救数据库表单信息**(表 2)：急救数据库是一个大型关联型数据库,其中包括 12 个包含患者真实诊疗信息的表单。表单之间通过带有“ID”后缀的字段进行关联。其中 PATIENT\_ID 代表患者加密后的唯一 ID,同一患者多次就诊所产生 PATIENT\_ID 相同;VISIT\_ID 与就诊人次相匹配,同一患者多次就诊所产生 VISIT\_ID 不同;EMG\_UNIQUE\_ID 为患者收入抢救单元救治后所产生的唯一 ID,同一患者多次收入抢救单元救治所产生的 EMG\_UNIQUE\_ID 不同;其他字段的命名与所包含的患者信息相关。

表 2 急救数据库表单名称及解释说明	
表单名称	解释说明
PATIENTS	患者基本信息,是否院内死亡以及死亡时间
TRIAGE_PATIENTS	分诊时间、级别、区域、去向等
	患者预检分诊信息
EMG_PATIENTS_VISIT	患者入、出抢救单元时间及去向
VITAL_SIGNS	患者生命体征信息
CHARTEVENTS	患者血气分析及心肌损伤标志物化验信息
MEDICAL_ORDER	医嘱及其执行时间相关信息
MEDICAL_RECORD	急诊病历、抢救记录、病程记录、离科小结等医疗记录
NURSING_RECORD	护理记录
LAB_TEST_MASTER	实验室检验项目及标本采集时间等信息
LAB_RESULT	实验室检验结果
MEDICAL_EXAMINATION	影像学、胃镜、肠镜等辅助检查结果
EMG_INOUT_RECORD	患者出入量记录

3 讨 论

解放军总医院急救数据库具有数据体量大、种类多、规范化程度高等特点。基于该急救数据库,能为临床人员和工程人员针对开展急救辅助决策支持、风险预测预警、急救流程优化、分级诊疗决策等相关临床科学研究提供数据资源。国内外关于数据库建立与管理的先进模式为急救数据库的建立提供了宝贵经验。2016 年,美国科研团队在其原有 MIMIC-Ⅱ数据库基础上进一步改进升级,成功建立了 MIMIC-Ⅲ数据库,包含 52 423 例次就诊于波士顿贝斯医疗中心各重症加强治疗病房(ICU)的危重症患者信息,也是目前世界上少数可免费使用的高质量数据库之一<sup>[9]</sup>。高质量

表 1 急救数据库患者信息种类	
信息种类	内容描述
基本信息	性别、出生日期、死亡日期及时间
就诊信息	分诊时间、入抢救单元时间、离科时间及去向
诊断信息	患者在抢救单元救治阶段诊断信息,具体诊断名称根据国际疾病分类 ICD-9 进行规范化处理
医嘱信息	使用药物名称、剂量、频次、途径、使用时间,气管插管、胸外心脏按压等重要抢救措施实施时间,连续性肾脏替代治疗(CRRT)等重要救治措施实施时间,动脉、静脉置管等有创操作位置及时间,输入血制品种类、剂量及时间等
生命体征信息	经护士确认或床旁监护仪自动记录的患者心率、呼吸频率、体温、收缩压、舒张压、氧饱和度数值及采集时间
出量信息	尿液、胃液、引流液等种类、量值及时间
实验室检验	血常规、血气分析、血生化、凝血功能、尿粪常规、微生物学等相关实验室检验结果及标本采集时间
其他辅助检查	影像学、胃镜、肠镜、支气管镜、病理等辅助检查报告
医疗记录	急诊病历、抢救记录、出科小结、护理记录等文本数据



数据库的建立为基于数据技术的科学研究提供了坚实基础,数据库中的数据种类、数据特点决定了基于该数据库所能开展的科学研究类型。MIMIC-III数据库所包含的数据来源于在各重症ICU接受治疗的患者,为进行重症医学领域的科学研究提供了条件<sup>[9]</sup>。当前解放军总医院急救数据库包含的数据主要来源于到达急救医疗机构后接受紧急救治的患者数据,可为研究该救治阶段的时效救治、紧急救治、流程优化等相关问题提供多病种数据。

在国家政策领导下,我国在数据领域呈现蓬勃向上的发展趋势,众多医疗机构为适应其科研需要已完成或初步完成了数据库建设,成果颇丰。例如,刘师伟等<sup>[10]</sup>研发的Idata数据库,可用于糖尿病领域的相关研究;成亮等<sup>[11]</sup>基于湖南省骨盆创伤患者信息建立了骨盆创伤患者数据库;顾大川等<sup>[12]</sup>成功建立了中国成人心血管外科注册登记数据库;人类特殊表达基因数据库的建立也为肿瘤领域相关研究提供了数据基础<sup>[13]</sup>。

我国在数据领域的发展仍处于起步阶段<sup>[14]</sup>,尚存在诸多瓶颈亟待解决。一是,在我国当前医疗体制下,各个医疗机构之间还未建立广泛的数据互联互通共享机制,“信息孤岛”现象仍然严重。以急诊医学为例,由于大多数院前救治机构与院内急救科室存在数据互联互通障碍,在患者到达医疗机构前,医护人员对患者情况通常“知之甚少”;另外,由于医疗机构间数据传输的限制,使得当前医疗机构所建立的数据库多为单中心数据库,在广泛人群中的代表性可能受到局限。二是,受制于标准数据单元和数据规范缺失<sup>[15]</sup>。例如,医生对患者的诊断并不规范,相同疾病的诊断名称可能五花八门,为数据检索、提取和交换共享带来困难。三是,不同医疗机构之间数据结构不同,也为建立多中心数据库带来较大障碍。四是,随着产生数据的渠道不断增多,数据的真实性难以判断,数据质量难以保证,进而可能影响数据库质量<sup>[16]</sup>。

#### 4 展 望

急救数据库的建立为研究与急诊医学相关的时效救治、紧急救治等科研问题提供了数据基础,库内数据准确性和规范化程度均较高,能满足临床科研需要。但目前,库内数据所包含的救治阶段信息仍较局限,未与院前救治及住院后患者数据对接形成更高价值的数据链,在该领域,陈意飞和张劲松<sup>[17]</sup>建立院前院内急诊信息共享系统,实现急救领域院前院内信息对接与信息采集,诸多类似研究为急救全流程数据库的建立创造了有利条件,这也是我们下一步希望着重开展的工作。

#### 参考文献

- [1] Walsh SH. The clinician's perspective on electronic health records and how they can affect patient care [J]. *BMJ*, 2004, 328 (7449): 1184-1187. DOI: 10.1136/bmj.328.7449.1184.
- [2] Ishikawa KB. Medical big data for research use: current status and related issues [J]. *Japan Med Assoc J*, 2016, 59 (2-3): 110-124.
- [3] 卢喜烈. 开启人工智能心电分析系统新时代 [J]. *实用心电学杂志*, 2018, 27 (1): 4-6. DOI: 10.13308/j.issn.2095-9354.2018.01.001.
- Lu XL. Open a new era of artificial intelligence electrocardiogram analysis system [J]. *J Pract Electrocardiol*, 2018, 27 (1): 4-6. DOI:

- 10.13308/j.issn.2095-9354.2018.01.001.
- [4] Phinney MA, Zhuang Y, Lander S, et al. Contrast mining for pattern discovery and descriptive analytics to tailor sub-groups of patients using big data solutions [J]. *Stud Health Technol Inform*, 2017, 245: 544-548. DOI: 10.3233/978-1-61499-830-3-544.
- [5] Binder H, Blettner M. Big data in medical science: a biostatistical view [J]. *Dtsch Arztebl Int*, 2015, 112 (9): 137-142. DOI: 10.3238/arztebl.2015.0137.
- [6] Joudaki H, Rashidian A, Minaei-Bidgoli B, et al. Using data mining to detect health care fraud and abuse: a review of literature [J]. *Glob J Health Sci*, 2014, 7 (1): 194-202. DOI: 10.5539/gjhs.v7n1p194.
- [7] Zimmerman AM, Marwaha J, Nunez H, et al. Preoperative myocardial injury as a predictor of mortality in emergency general surgery: an analysis using the American College of Surgeons NSQIP Database [J]. *J Am Coll Surg*, 2016, 223 (2): 381-386. DOI: 10.1016/j.jamcollsurg.2016.04.043.
- [8] 段立伟, 张晟, 林兆奋. 以 logistic 回归模型构建联合预测因子对脓毒症诊断及预后判断的临床运用 [J]. *中华危重病急救医学*, 2017, 29 (2): 139-144. DOI: 10.3760/cma.j.issn.2095-4352.2017.02.009.
- Duan LW, Zhang S, Lin ZF. Formulation of combined predictive indicators using logistic regression model in predicting sepsis and prognosis [J]. *Chin Crit Care Med*, 2017, 29 (2): 139-144. DOI: 10.3760/cma.j.issn.2095-4352.2017.02.009.
- [9] Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database [J]. *Scientific data*, 2016 (35): 1-9. DOI: 10.1038/sdata.2016.35.
- [10] 刘师伟, 段瑞雪, 李欣, 等. 医学大数据库的建立与哲学思辨 [J]. *医学与哲学*, 2016, 37 (21): 10-13. DOI: 10.12014/j.issn.1002-0772.2016.11a.03.
- Liu SW, Duan RX, Li X, et al. Critical thinking of medical big databases establishment and its philosophical reflection [J]. *Med Philosophy*, 2016, 37 (21): 10-13. DOI: 10.12014/j.issn.1002-0772.2016.11a.03.
- [11] 成亮, 朱勇, 龙海涛, 等. 湖南省区域性骨盆创伤数据库的研制 [J]. *中南大学学报 (医学版)*, 2017, 42 (4): 445-450. DOI: 10.11817/j.issn.1672-7347.2017.04.013.
- Cheng L, Zhu Y, Long HT, et al. Establishment for regional pelvic trauma database in Hunan Province [J]. *J Centr South Univ Med Sci*, 2017, 42 (4): 445-450. DOI: 10.11817/j.issn.1672-7347.2017.04.013.
- [12] 顾大川, 郑哲, 赵韡, 等. 中国成人心血管外科注册登记数据库的构建 [J]. *中国循环杂志*, 2017, 32 (10): 1010-1014. DOI: 10.3969/j.issn.1000-3614.2017.10.017.
- Gu DC, Zheng Z, Zhao W, et al. Establishment of Chinese Cardiovascular Surgery Registry Database [J]. *Chin Circ J*, 2017, 32 (10): 1010-1014. DOI: 10.3969/j.issn.1000-3614.2017.10.017.
- [13] Tang Q, Zhang Q, Lyu Y, et al. SEGREG: a database for human specifically expressed genes and their regulations in cancer and normal tissue [J/OL]. *Brief Bioinform*, 2018 [2018-01-05]. [published online ahead of print January 3, 2018]. DOI: 10.1093/bib/bbx173.
- [14] 王潇, 张爱迪, 严谨. 大数据在医疗卫生中的应用前景 [J]. *中国全科医学*, 2015, 16 (1): 113-115. DOI: 10.3969/j.issn.1007-9572.2015.01.028.
- Wang X, Zhang AD, Yan J. Application prospects of big data in health care [J]. *Chin Gen Pract*, 2015, 16 (1): 113-115. DOI: 10.3969/j.issn.1007-9572.2015.01.028.
- [15] 林敏. 健康医疗大数据的应用与发展 [J]. *医疗装备*, 2017, 30 (1): 198-199. DOI: 10.3969/j.issn.1002-2376.2017.01.138.
- Lin M. Applications and development of health big data [J]. *Med Equipment*, 2017, 30 (1): 198-199. DOI: 10.3969/j.issn.1002-2376.2017.01.138.
- [16] 沈韬, 崔泳. 医疗大数据: 期望与现实 [J]. *中国数字医学*, 2015, 10 (7): 2-4, 32. DOI: 10.3969/j.issn.1673-7571.2015.07.001.
- Shen T, Cui Y. Expectation and reality for medical big data [J]. *China Dig Med*, 2015, 10 (7): 2-4, 32. DOI: 10.3969/j.issn.1673-7571.2015.07.001.
- [17] 陈意飞, 张劲松. 院前院内急诊信息共享系统的构建与探讨 [J]. *中华危重病急救医学*, 2017, 29 (3): 268-269. DOI: 10.3760/cma.j.issn.2095-4352.2017.03.015.
- Chen YF, Zhang JS. Construction and outlook of sharing information in pre-hospital emergency and emergency disposal in hospital [J]. *Chin Crit Care Med*, 2017, 29 (3): 268-269. DOI: 10.3760/cma.j.issn.2095-4352.2017.03.015.

(收稿日期: 2018-02-26)