·论著·

危急重症监护数据库MIMIC-Ⅲ疾病谱分析

范勇 赵宇卓 李沛尧 刘晓莉 贾立静 李开源 冯聪 潘菲 黎檀实 张政波 曹德森

100853 北京,解放军总医院医学工程保障中心(范勇、李沛尧、张政波、曹德森),急诊科(赵宇卓、贾立静、李开源、冯聪、潘菲、黎檀实),医疗大数据中心(张政波);100191 北京航空航天大学生物与医学工程学院(刘晓莉)

通讯作者:张政波,Email:zhengbozhang@126.com

DOI:10.3760/cma.j.issn.2095-4352.2018.06.006

【摘要】 目的 深度解析重症监护医学信息数据库Ⅲ(MIMIC-Ⅲ)患者疾病谱,为基于MIMIC-Ⅲ数据库解决临床科研问题的临床医生及工程师提供相关数据参考。**方法** 利用探索性数据分析方法,探究MIMIC-Ⅲ数据库中各年龄层患者(不包括新生儿)疾病及急症分布特点;基于相同方法,分析新生儿孕周、体重、重症加强治疗病房(ICU)住院时间等数据的分布特点。**结果** MIMIC-Ⅲ数据库中首次入院46 428例患者,49 214例次ICU记录。其中男性26 076例,女性20 352例;中位年龄为60.5(38.6,75.6)岁;分布在61~80岁的患者最多。疾病谱分析中,第一诊断以循环系统疾病患者占比最大(占32%),其次为损伤和中毒(占14%)、消化系统疾病(占8%)、肿瘤(占7%)、呼吸系统疾病(占6%)等。循环系统疾病中缺血性心脏病患者占比最大(占42%),患者比例随年龄增加到60~70岁达最大值后逐渐下降;而脑血管疾病患者比例则随年龄增长呈先下降后升高趋势,并且是循环系统疾病死亡的主要原因(占22.5%)。损伤和中毒患者随年龄增加比例呈明显下降趋势。消化系统疾病较总人群分布偏年轻化(50~60岁者最多),非感染性肠炎和结肠炎是其主要死亡原因(ICU病死率18.3%)。在感染患者中以呼吸系统感染为主(占34%),但循环系统感染是其主要死亡原因(ICU病死率25.6%)。监护室新生儿中早产儿占82%,随孕龄增加,ICU住院时间减少,且病死率下降。**结论** 通过对MIMIC-Ⅲ数据库患者疾病谱进行深度解析,能为相关领域研究者提供一定数据参考,利于先期掌握研究目标对象的体量和分布概况以及开展下一步研究,同时可了解探索性数据分析技术在医疗数据分析领域的重要作用,为利用电子健康档案进行数据研究提供便利。

【关键词】 疾病谱; 危急重症; 重症监护医学信息数据库Ⅲ; 电子健康档案; 探索性数据分析

基金项目:国家自然科学基金(61471398);军队保健专项(16BJZ23);解放军总医院转化医学项目(2016TM-041);医疗大数据应用技术国家工程实验室(2017-148)

Analysis of diseases distribution in Medical Information Mart for Intensive Care Ⅲ database

Fan Yong, Zhao Yuzhuo, Li Peiyao, Liu Xiaoli, Jia Lijing, Li Kaiyuan, Feng Cong, Pan Fei, Li Tanshi, Zhang Zhengbo, Cao Desen

Department of Biomedical Engineering and Maintenance Center, Chinese PLA General Hospital, Beijing 100853, China (Fan Y, Li PY, Zhang ZB, Cao DS); Department of Emergency, Chinese PLA General Hospital, Beijing 100853, China (Zhao YZ, Jia LJ, Li KY, Feng C, Pan F, Li TS); Medical Information Center, Chinese PLA General Hospital, Beijing 100853, China (Zhang ZB); School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China (Liu XL)

Corresponding author: Zhang Zhengbo, Email: zhengbozhang@126.com

【Abstract】 Objective To study the distribution of diseases in Medical Information Mart for Intensive Care Ⅲ (MIMIC-Ⅲ) database in order to provide reference for clinicians and engineers who use MIMIC-Ⅲ database to solve clinical research problems. **Methods** The exploratory data analysis technologies were used to explore the distribution characteristics of diseases and emergencies of patients (excluding newborns) in MIMIC-Ⅲ database were explored; then, neonatal gestational age, weight, length of hospital stay in intensive care unit (ICU) were analyzed with the same method. **Results** In the MIMIC-Ⅲ database, 46 428 patients were admitted for the first time, and 49 214 ICU records were recorded. There were 26 076 males and 20 352 females; the median age was 60.5 (38.6, 75.6) years, and most patients were between 60 and 80 years old. The first diagnosis in the disease spectrum analysis was firstly ranked by circulatory diseases (32%), followed by injury and poisoning (14%), digestive system disease (8%), tumor (7%), respiratory disease (6%) and so on. Patients with ischemic heart disease accounted for the largest proportion of circulatory disease (42%), the proportion of these patients gradually increased with age of 60-70 years old, then decreased. However, the proportion of patients with cerebrovascular disease declined first and then increased with age, which was the main cause of death of circulatory system disease (ICU mortality was 22.5%). Injury and poisoning patients showed a significant decrease with age. Digestive system diseases were younger than the general population (most people aged between 50 to 60 years), and non-infectious enteritis and colitis were the main causes of death (ICU mortality was 18.3%). Respiratory infections were predominant in infected patients (34%), but circulatory system infections were the main cause of death (ICU mortality was 25.6%). Secondly, in the neonatal care unit, premature infants accounted for the vast majority (82%). As the gestational age increased, the duration of ICU was decreased, and the mortality was

decreased. **Conclusions** The diseases distribution of patients can be provided by MIMIC-Ⅲ database, which helps to grasp the overview of the volume and age distribution of the target patients in advance, and carry out the next step of research. Meanwhile, it points out the important role of exploratory data analysis in electronic health records analysis.

【Key words】 Spectrum of disease; Critical care; Medical Information Mart for Intensive Care Ⅲ; Electronic health records; Exploratory data analysis

Fund program: National Natural Science Foundation of China (61471398); Military Health Care Program (16BJZ23); Chinese PLA General Hospital Transformation Medical Project (2016TM-041); National Engineering Laboratory for Big-data Application Technology (2017-148)

随着医疗信息化建设的高速发展，医疗领域已经积累了大量的电子健康档案（EHR）数据，包括：来自各级医院的医疗记录，如生命体征信息、实验室检查、影像学检查、基因数据；公共卫生服务机构数据；地方卫生局行政管理数据等等[1]，其中既有大量结构化数据，也有非结构化数据。对 EHR 分析利用一直是医务工作者及其他相关研究人员关注的重点，包括循证医学、公共卫生领域、药物研究开发、基因分析等各方面的研究[2]。

以重症监护医学信息数据库（MIMIC）为例，它是由麻省理工学院计算生理实验室建立的大样本、单中心危急重症监护数据库，包含了美国波士顿 BID 医学中心（Beth Israel Deaconess Medical Center）重症加强治疗病房（ICU）去隐私化的医疗记录，并免费提供给全球研究者进行学术研究；其数据类型包括患者生命体征、实验室检查结果、药物使用、护理记录、手术操作代码、疾病诊断代码等。最新版本 MIMIC-Ⅲ于 2015 年年底发布，包含了 49 785 例患者的入院记录，以及从 2001 至 2012 年 53 423 例次年龄≥16 岁的 ICU 患者记录。相比 MIMIC-Ⅱ数据库，MIMIC-Ⅲ数据库增加了 2.8 万条记录，而且在数据清洗校对方面做了更多工作，使其结构更加简单，数据可靠度更高[3]。目前基于 MIMIC-Ⅲ也已开展了多项研究，如：脓毒症发生时间预测研究[4]、患者结局预测研究[5]等。

机器学习、大数据技术在医学上成功应用的案例，使越来越多的研究者和临床医生对利用医疗数据资源进行研究产生了极大的兴趣。目前国内外基于医疗数据库开展临床科学研究的思路通常是进行回顾性研究，首先由临床医生根据工作中需解决的重点难点问题提出需求，再与工程师组成跨学科团队，基于 EHR 共同解决问题。在该种模式中的重点环节之一就在于由临床医生提出合理化的科学问题，工程师围绕该核心问题进行数据提取、建模和分析[6]。然而对于从事一线临床工作的医生和学者来说，由于缺乏 SQL 编程和 EHR 数据库架构的相关知识[7]，对 MIMIC-Ⅲ等大型数据库疾病谱认知的缺失，其提出问题的模式还基于自身所在医疗机构疾病谱特点，导致所提出的临床问题得不到数据支撑。因此，让临床医生或研究者先期了解数据库中的数据内容是促进其与工程师交流合作，高效利用 EHR 进行二次分析的重要手段。本研究通过探索性数据分析的方法，旨在将 MIMIC-Ⅲ数据库患者疾病谱详尽呈现，从而为利用 MIMIC-Ⅲ进行科学研究提供数据基础及参考。

1　资料与方法

1.1　数据来源：数据来源于 MIMIC-Ⅲ（v1.4）数据库，该版本数据库共包含 46 520 例患者、61 532 例次 ICU 诊疗记录。本研究主要通过 PostgreSQL 9.6 软件提取数据库中患者人口统计学基本信息、诊断和结局等数据。

1.2　研究方法：首先，呈现 MIMIC-Ⅲ数据库中患者部分基本信息。其次，根据疾病诊断进行探索性数据分析。鉴于新生儿与成人间的巨大差异，新生儿部分单独分析。对于非新生儿患者，以第一诊断分析各疾病类型在不同年龄段之间的分布，并按照国际疾病分类代码 ICD-9[8]逐步细化疾病分类，分析几大类主要疾病类型在各年龄段之间的基本信息。首先按照 Angus 定义感染的 ICD-9 代码来确定感染人群，然后根据发生全身炎症反应综合征（SIRS）来确定脓毒症，根据发生器官衰竭确定为严重脓毒症[9]。对于新生儿部分，主要从新生儿的孕周、体重、ICU 时长等特征进行分析，并对死亡案例进行分析，初步了解 ICU 内新生儿死亡特征。以上分析均是基于患者首次入院数据，再入院情况暂时不在考虑范围。

1.3　统计学方法：正态分布的连续变量以均数 ± 标准差（$\bar{x} \pm s$）表示，采用 t 检验；非正态分布的连续变量以中位数（四分位数）〔$M(Q_L, Q_U)$〕表示，采用 Kruskal-Wallis 检验。分类变量使用 χ^2 检验。数据分析在统计语言 R 中完成（R-3.3.2）[10]。

2　结　果

2.1　MIMIC-Ⅲ数据库基本情况：诊疗数据来源于内科 ICU（MICU）、心脏外科 ICU（CSRU）、外科 ICU

表 1 重症监护医学信息数据库Ⅲ（MIMIC-Ⅲ）中首次入院患者的基本信息

科室	患者数〔例（%）〕	ICU 住院数（例次）	年龄〔岁，$M(Q_L, Q_U)$〕	男性（%）	住院时间〔d，$M(Q_L, Q_U)$〕	ICU 住院时间〔d，$M(Q_L, Q_U)$〕	院内病死率（%）
MICU	13 610（ 29.3）	14 500	64.8（50.7, 77.4）	51.1	6.0（4.0, 12.0）	2（1, 4）	15.0
CSRU	7 605（ 16.4）	8 157	67.9（58.2, 76.6）	66.4	7.0（5.0, 11.0）	2（1, 3）	3.8
SICU	6 362（ 13.7）	6 959	64.0（51.5, 77.3）	51.4	8.0（4.0, 14.0）	2（1, 5）	13.2
CCU	5 692（ 12.3）	6 009	70.7（58.3, 82.8）	58.1	5.0（3.0, 10.0）	2（1, 4）	11.5
TSICU	5 297（ 11.4）	5 619	70.6（58.5, 80.6）	61.5	7.0（4.0, 14.0）	2（1, 5）	11.3
NICU	7 862（ 16.9）	7 970	0.0（ 0.0, 0.0）	54.0	4.0（2.0, 10.0）	1（0, 9）	0.7
全部	46 428（100.0）	49 214	60.5（38.6, 75.6）	56.2	6.0（4.0, 12.0）	2（1, 4）	9.7

注：MICU 为内科重症加强治疗病房，CSRU 为心脏外科重症加强治疗病房，SICU 为外科重症加强治疗病房，CCU 为心内科重症加强治疗病房，TSICU 为创伤外科重症加强治疗病房，NICU 为新生儿重症加强治疗病房，ICU 为重症加强治疗病房

（SICU）、心内科 ICU（CCU）、创伤外科 ICU（TSICU）和新生儿 ICU（NICU），其中包括患者人口统计学信息、实验室化验结果、药物使用记录、诊疗记录和院外死亡相关信息、生理波形等共计 26 张表，如：ADMISSIONS、CHARTEVENTS、D-ITEMS、DIAGNOSES_ICD 等。

本次纳入首次入院 46 428 例患者、49 214 例次 ICU 记录进行分析（表 1），不包含再入院 6 555 例患者、11 837 例次 ICU 记录，并舍弃了基本信息不完整者。图 1 显示，患者多集中在 50 岁以上，中位年龄为 60.5（38.6，75.6）岁，表明 ICU 中以老年人居多。

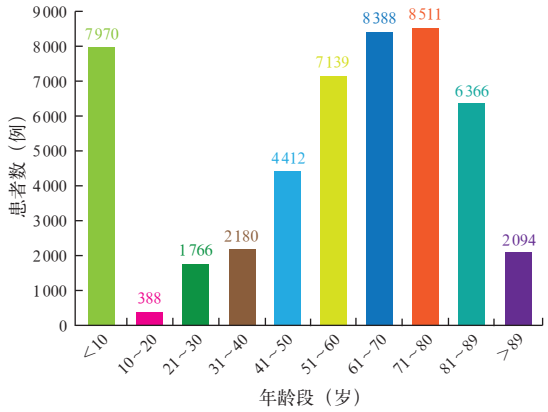


图 1 重症监护医学信息数据库Ⅲ（MIMIC-Ⅲ）中首次入院患者的年龄分布

患者以白种人为主，约占 70%，其次是黑种人、黄种人等（图 2）。患者第一疾病诊断以循环系统疾病居多，其次为损伤和中毒、消化系统疾病、肿瘤、呼吸系统疾病、传染病和寄生虫病等（图 3）。

2.2 MIMIC-Ⅲ患者诊断探索性数据分析：各年龄段患者第一诊断分布见图 4。多数疾病均随年龄增长有明显的改变趋势，例如：循环系统疾病患者比例随年龄的增长逐渐增加，在 71～80 岁年龄段达到最高；损伤和中毒类患者比例随年龄的增长逐渐减少；而呼吸系统、消化系统疾病患者在各年龄段所占比例差异不大。

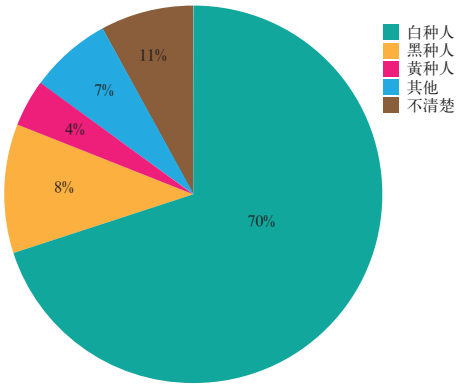


图 2 重症监护医学信息数据库Ⅲ（MIMIC-Ⅲ）中首次入院患者的种族分布

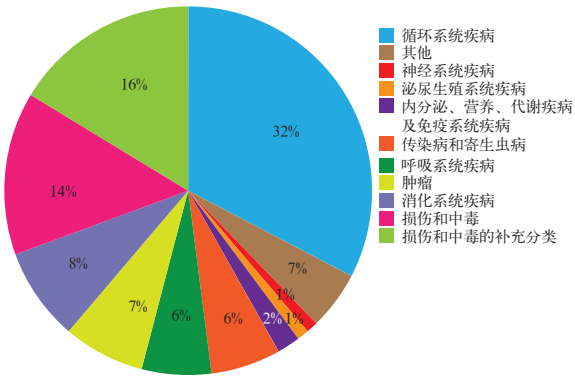


图 3 重症监护医学信息数据库Ⅲ（MIMIC-Ⅲ）中首次入院患者第一诊断分布

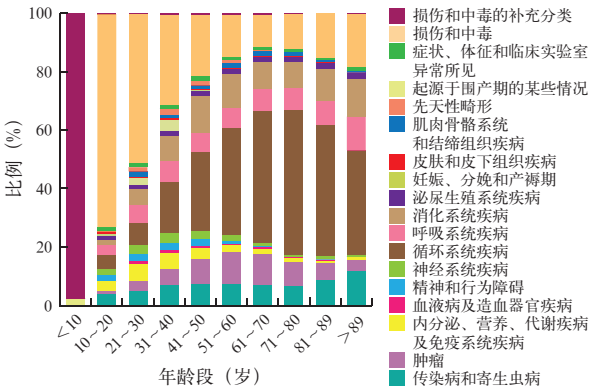


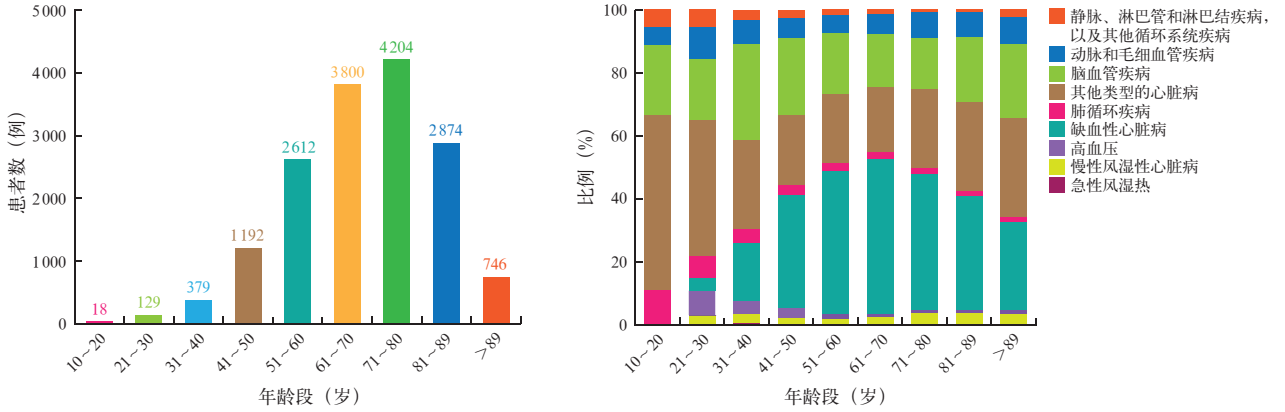
图 4 重症监护医学信息数据库Ⅲ（MIMIC-Ⅲ）中不同年龄段患者第一诊断分布

按 ICD-9 代码进一步细化循环系统、消化系统、呼吸系统疾病分类。

2.2.1 循环系统疾病（图 5；表 2）：共有 15 954 例首次入 ICU 患者记录，各年龄段分布与总人群分布保持一致，71～80 岁患者最多；且同样可以看到疾病随年龄变化的趋势，如：局部缺血性心脏病患者比例随年龄增加到 61～70 岁达最大值后逐渐下降；而脑血管疾病患者比例随年龄增长呈先下降后升高的趋势，在 61～70 岁年龄段中所占比例最小。循环系统疾病中，缺血性心脏病患者占比最大，其中男性占 70.8%；其次为脑血管疾病，也是导致 ICU 患者死亡最常见的循环系统疾病，ICU 病死率高达 22.5%。

2.2.2 消化系统疾病（图 6；表 3）：共有 4 163 例首次入 ICU 患者记录，其疾病分布在 51～60 岁患者最多，较总人群分布偏年轻化；进一步细分显示，各消化系统疾病与年龄分布没有明显关系。不同消化系统疾病中非感染性小肠炎和结肠炎患者死亡比例最高。

2.2.3 呼吸系统疾病（图 7；表 4）：共有 3 117 例首次入 ICU 患者记录，其疾病分布与总人群分布一致；进一步细分显示，流行性感冒和肺炎患者最多，与临



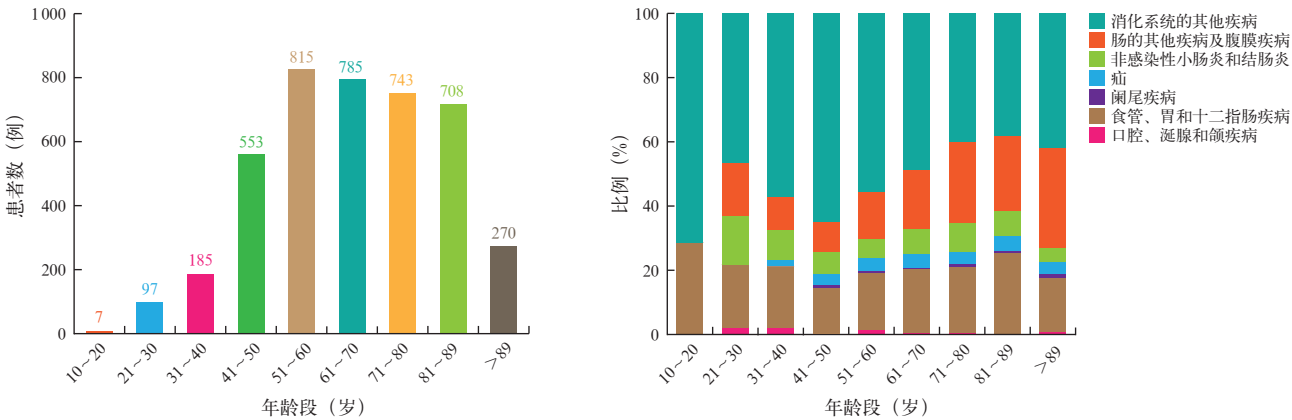
注：ICU 为重症加强治疗病房

图 5 重症监护医学信息数据库Ⅲ（MIMIC-Ⅲ）中首次入 ICU 循环系统疾病患者不同年龄段分布（左）及第一诊断分布（右）

表 2 重症监护医学信息数据库Ⅲ（MIMIC-Ⅲ）首次入院的各类循环系统疾病患者基本信息

循环系统疾病	患者数（例）	ICU 住院数（例次）	年龄〔岁，$M(Q_L, Q_U)$〕	男性（%）	住院时间〔d，$M(Q_L, Q_U)$〕	ICU 住院时间〔d，$M(Q_L, Q_U)$〕	ICU 病死率（%）
缺血性心脏病	6 314	6 640	68.6（59.6，77.5）	70.8	6.0（4.0，10.0）	2（1，3）	4.8
脑血管疾病	2 865	3 020	69.3（56.6，80.2）	49.6	6.0（3.0，12.0）	2（1，5）	22.5
动脉和毛细血管疾病	1 045	1 130	71.1（60.3，79.8）	56.6	9.0（5.0，17.0）	3（1，7）	15.0
慢性风湿性心脏病	457	495	74.7（63.0，81.4）	44.2	9.0（6.0，14.0）	2（1，5）	5.5
其他类型的心脏病	3 685	3 909	71.5（59.5，80.6）	57.4	7.0（4.0，10.0）	2（1，4）	7.1

注：ICU 为重症加强治疗病房



注：ICU 为重症加强治疗病房

图 6 重症监护医学信息数据库Ⅲ（MIMIC-Ⅲ）中首次入 ICU 消化系统疾病患者不同年龄段分布（左）及第一诊断分布（右）

床一般认知相同。外部物质引起的肺部疾病是引起呼吸系统疾病患者死亡的最主要因素。

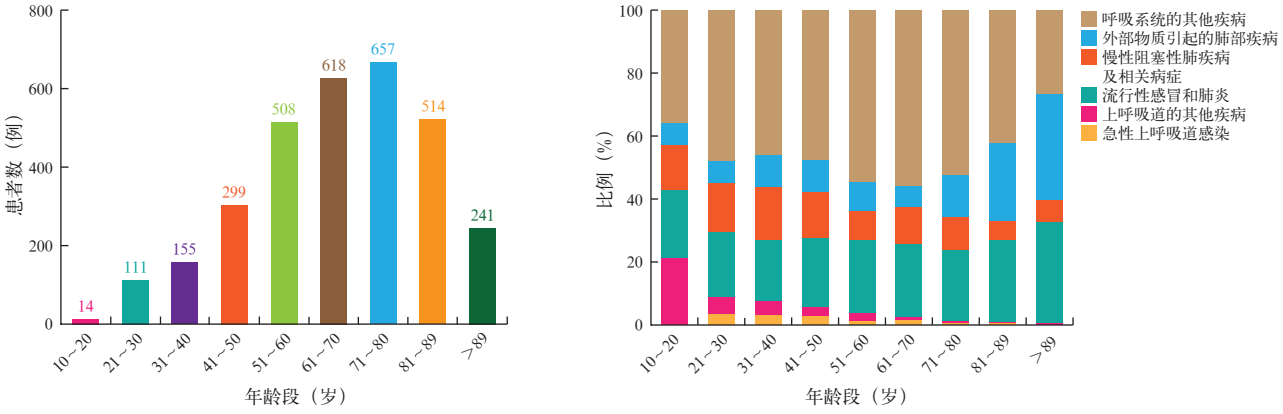
2.2.4 感染（图 8；表 5）：共有 15 453 例感染患者、17 344 例次 ICU 诊疗记录。呼吸系统感染患者最多，与临床一般认知相符合。但循环系统感染是引起感染患者死亡的最主要原因。

2.2.5 脓毒症（表 6）：在 MIMIC-Ⅲ数据库中，脓毒症患者 ICU 病死率高，多集中在内科 ICU。

表 3 重症监护医学信息数据库Ⅲ（MIMIC-Ⅲ）中首次入院的各类消化系统疾病患者基本信息

消化系统疾病	患者数（例）	ICU 住院数（例次）	年龄〔岁，$M(Q_L, Q_U)$〕	男性（%）	住院时间〔d，$M(Q_L, Q_U)$〕	ICU 住院时间〔d，$M(Q_L, Q_U)$〕	ICU 病死率（%）
食管、胃和十二指肠疾病	766	813	71.5（59.5，80.6）	57.4	7.0（4.0，10.0）	2（1，4）	7.1
肠的其他疾病及腹膜疾病	732	785	73.9（59.6，82.5）	45.6	8.0（5.0，14.0）	2（1，4）	8.6
非感染性小肠炎和结肠炎	267	312	66.6（50.9，77.3）	44.9	9.0（5.5，18.0）	2（1，6）	18.3
疝	145	166	67.1（55.5，80.1）	43.4	10.0（6.0，17.0）	2（1，6）	4.8
消化系统的其他疾病	1 822	2 028	61.7（50.3，77.0）	57.2	8.0（4.0，10.0）	2（1，4）	13.5

注：ICU 为重症加强治疗病房



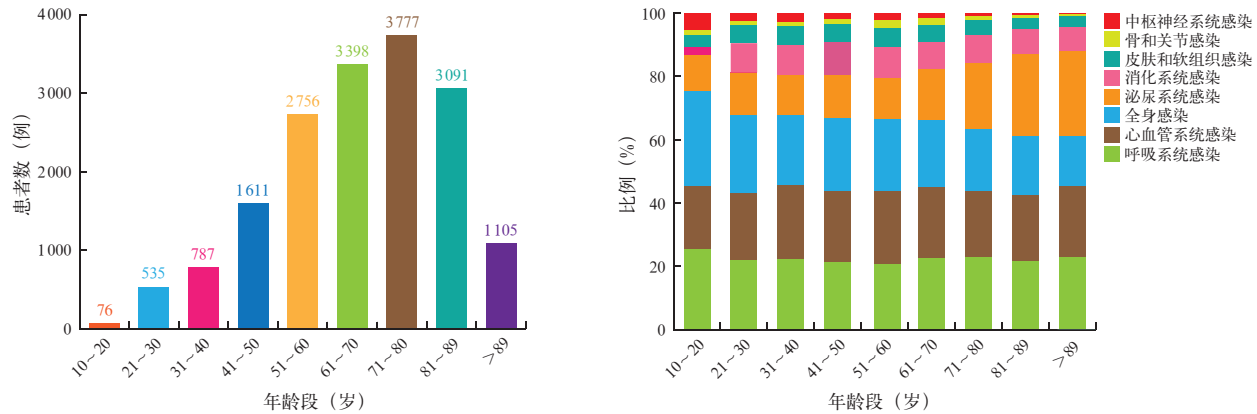
注：ICU 为重症加强治疗病房

图 7 重症监护医学信息数据库Ⅲ（MIMIC-Ⅲ）中首次入 ICU 呼吸系统疾病患者不同年龄段分布（左）及第一诊断分布（右）

表 4 重症监护医学信息数据库Ⅲ（MIMIC-Ⅲ）中首次入院的各类呼吸系统疾病患者基本信息

呼吸系统疾病	患者数（例）	ICU 住院数（例次）	年龄〔岁，$M(Q_L, Q_U)$〕	男性（%）	住院时间〔d，$M(Q_L, Q_U)$〕	ICU 住院时间〔d，$M(Q_L, Q_U)$〕	ICU 病死率（%）
流行性感冒和肺炎	696	744	69.2（55.6，81.7）	51.1	8.0（5.0，12.2）	3（1，6）	18.2
外部物质引起的肺部疾病	408	439	78.1（61.1，86.5）	52.2	8.0（5.0，12.0）	2（1，5）	18.6
慢性阻塞性肺疾病及相关病症	316	325	64.1（48.2，75.4）	38.0	5.0（3.0，9.0）	2（1，4）	5.7
呼吸系统的其他疾病	1 399	1 517	66.3（53.7，77.5）	50.7	8.0（4.0，10.0）	4（2，8）	21.4

注：ICU 为重症加强治疗病房



注：ICU 为重症加强治疗病房

图 8 重症监护医疗数据库Ⅲ（MIMIC-Ⅲ）中首次入 ICU 感染患者年龄分布（左）及第一诊断分布（右）

2.3 MIMIC-Ⅲ新生儿探索性数据分析(图 9;表 7):选取孕龄和 ICU 住院时间完整的新生儿记录,根据新生儿基本特征表明,引起新生儿死亡的因素与孕龄短、出生体重轻、身高矮等相关。

MIMIC 数据库中,早产儿 2 744 例中死亡 41 例,足月产 612 例中死亡 1 例;病死率随孕龄增加逐渐下降,孕龄在 25 周的新生儿病死率最高,达到 22%。新生儿 ICU 住院时间随孕龄增加逐渐减少。

3　讨　论

本研究基于最新版本的多参数重症监护数据库 MIMIC-Ⅲ对首次入院患者基本信息以及疾病谱进行了探索性数据分析,提供了不同疾病患者的体量及分布概况,为基于 MIMIC-Ⅲ数据库解决临床科研问题的临床医生及工程师提供相关数据参考。例如:Lee 等[7]曾基于 MIMIC-Ⅱ数据库开发了网页版的数据化可视工具,帮助对 MIMIC-Ⅱ感兴趣的研究者

表 5　重症监护医学信息数据库Ⅲ(MIMIC-Ⅲ)中首次入院的各类感染患者基本信息

感染类型	患者数(例)	ICU 住院数(例次)	年龄〔岁,$M(Q_L, Q_U)$〕	男性(%)	住院时间〔d,$M(Q_L, Q_U)$〕	ICU 住院时间〔d,$M(Q_L, Q_U)$〕	ICU 病死率(%)
呼吸系统感染	5 903	6 649	68.4(54.6, 79.5)	55.1	11.0(6.0, 20.0)	4(2, 10)	18.7
心血管系统感染	5 735	6 558	67.0(54.3, 78.6)	56.9	11.0(6.0, 21.0)	3(2, 9)	25.6
全身感染	5 329	6 221	66.0(52.3, 78.1)	49.8	14.0(8.0, 24.0)	3(2, 9)	11.7
泌尿系统感染	4 888	5 488	74.2(60.4, 83.1)	36.0	10.0(6.0, 17.0)	3(1, 6)	12.0
消化系统感染	2 171	2 623	66.8(53.7, 78.6)	52.7	14.0(7.0, 25.0)	3(2, 8)	20.0
皮肤和软组织感染	1 335	1 520	63.5(51.3, 75.9)	56.9	12.0(7.0, 21.0)	3(1, 6)	9.5
骨和关节感染	433	508	62.3(54.1, 73.8)	63.0	15.0(9.0, 25.0)	2(1, 6)	12.2
中枢神经系统感染	340	406	53.6(39.8, 63.9)	54.7	17.0(9.0, 30.0)	5(2, 14)	11.5

注:ICU 为重症加强治疗病房

表 6　重症监护医学信息数据库Ⅲ(MIMIC-Ⅲ)中首次入院的脓毒症患者基本信息

患者	例数(例)	年龄〔岁,$M(Q_L, Q_U)$〕	男性〔例(%)〕	ICU 住院时间〔d,$M(Q_L, Q_U)$〕	ICU 病死率〔%(例)〕	住院时间〔d,$M(Q_L, Q_U)$〕	住院病死率〔%(例)〕
非脓毒症	36 919	57.9(27.2, 73.5)	21 117(57.2)	2(1, 3)	5.3(1 952)	6.0(3.0, 10.0)	6.2(2 291)
脓毒症	12 295	68.5(54.9, 79.9)[a]	6 587(53.6)[a]	4(2, 9)[a]	16.7(2 056)[a]	13.0(7.0, 24.0)[a]	22.3(2 747)[a]

患者	例数(例)	入 ICU 类型分布〔例(%)〕[a]					
		CCU	CSRU	MICU	NICU	SICU	TSICU
非脓毒症	36 919	4 462(12.1)	7 087(19.2)	8 228(22.3)	7 836(21.2)	5 032(13.6)	4 274(11.6)
脓毒症	12 295	1 547(12.6)	1 070(8.7)	6 272(51.0)	134(1.1)	1 927(15.7)	1 345(10.9)

注:ICU 为重症加强治疗病房,CCU 为心内科重症加强治疗病房,CSRU 为心脏外科重症加强治疗病房,MICU 为内科重症加强治疗病房,NICU 为新生儿重症加强治疗病房,SICU 为外科重症加强治疗病房,TSICU 为创伤外科重症加强治疗病房;与非脓毒症患者比较,[a]$P<0.01$

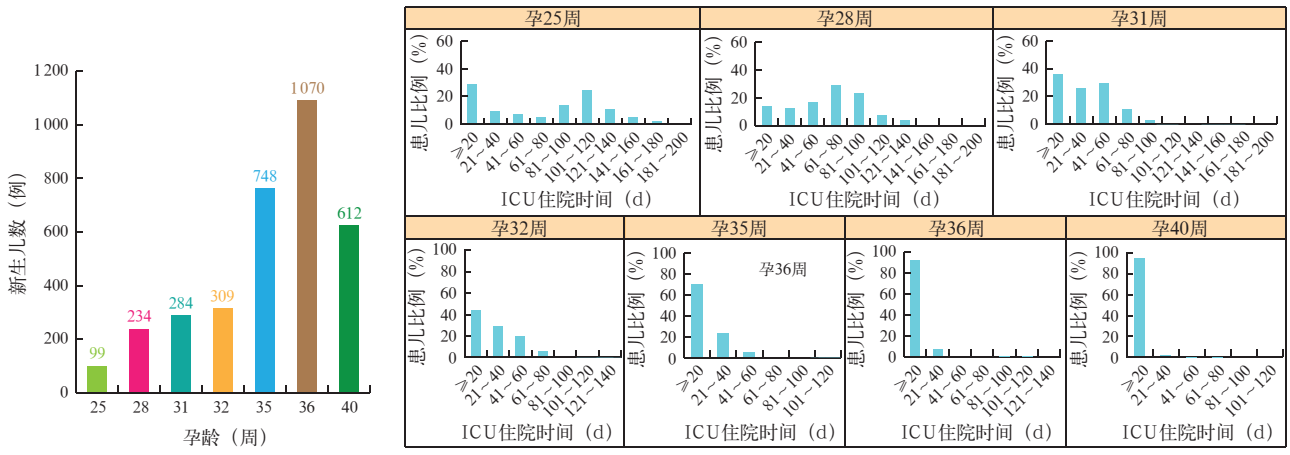


图 9　重症监护医学信息数据库Ⅲ(MIMIC-Ⅲ)中首次入院的 0 ~ 1 d 新生儿孕龄分布(左)及 ICU 住院时间分布(右)

表 7　重症监护医学信息数据库Ⅲ(MIMIC-Ⅲ)中首次入院的 0 ~ 1 d 新生儿基本信息

预后	患者数(例)	男性〔例(%)〕	孕龄〔周,$M(Q_L, Q_U)$〕	体重〔kg,$M(Q_L, Q_U)$〕	身高〔cm,$M(Q_L, Q_U)$〕	ICU 住院时间〔d,$M(Q_L, Q_U)$〕
存活	3 314	1 819(54.9)	36(32, 36)	1.8 (0.9, 2.6)	44.5(41.0, 47.0)	10(2, 24)
死亡	42	23(54.8)	25(25, 28)	0.66(0.4, 1.0)	33.0(31.5, 38.0)	2(25, 28)

注:ICU 为重症加强治疗病房

来了解其中患者的特性，从人口统计学、入院信息、患者结局、生理信号及实验室检查结果等多个方面进行了探索性数据分析，给刚接触使用 MIMIC-Ⅱ数据库的研究人员提供了帮助，虽然涉及面广，但不够细致，在目标人群的确立，给临床医生的参考价值较小。而 MIMIC-Ⅲ数据库中的患者数量大幅增加，数据质量也得到了进一步的保证。本研究主要针对不同年龄段患者疾病谱分布进行了详细描述，系统或大类疾病排序中前 5 位依次为循环系统、损伤和中毒、消化系统、肿瘤、呼吸系统，与我国大多数医院危急重症疾病谱相似，但不尽相同，国内排在前 5 位的一般为损伤和中毒、神经系统、消化系统、循环系统疾病、呼吸系统疾病[11-14]。同时针对各系统疾病中病死率较高病种，如：循环系统中的脑血管疾病，消化系统中的非感染性小肠炎和结肠炎以及呼吸系统中的流行性感冒和肺炎，需要加大认识和研究，加强相关的预防保健宣传、危险因素控制工作。

本研究不足之处：对新生儿只展示了其基本信息，未对其诊断进行分析，这项工作需要从新生儿出院诊断文本中获取信息。国内有研究表明，新生儿出院第一诊断排名前 3 位的疾病分别为呼吸系统疾病、高危产和新生儿黄疸[15]。对重症新生儿感兴趣的研究者可以依据 MIMIC-Ⅲ数据库进行儿科临床诊疗以及预防保健方面的相关研究。

探索性数据分析是数据科学中不可缺少的一步，其包含的方法多是将数据以图形的方式呈现出来，便于直观理解数据，发掘更深层次的关系、模式等。国内解放军总医院急诊科发布的急救数据库，收集了 2014 年 1 月至 2018 年 1 月急诊科就诊及收治的患者信息，包含了 50 万余例次分诊信息以及 2 万余例次急诊抢救单元收治患者的诊疗信息。在利用这些数据之前进行相应的探索性数据分析，能让研究者更加了解数据库，对开展临床研究起到促进作用。目前为了使用不同来源的 EHR 数据，许多研究者正致力于将以往的观察性数据库转化为观察性医疗结果通用数据模型（OMOP-CDM）[16]，MIMIC-Ⅲ的研究团队也在进行此项工作，基于通用数据模型中的数据和临床医生探讨研发出一些通用的可视化模型，可以大大节省前期研究者们熟悉数据过程的时间。

综上，本研究对 MIMIC-Ⅲ数据库患者疾病谱进行了详细的描述，提供了不同疾病患者的体量及分布概况，为临床医生利用临床大数据开展医学领域的“真实世界”研究进行了系统性的铺垫工作，打下了坚实的基础。未来，在此基础上可以进一步与临床医生合作开展基于 MIMIC-Ⅲ数据库的临床研究，从而获得有意义的循证医学证据。

参考文献

[1] Zhang L, Wang H, Li Q, et al. Big data and medical research in China [J]. BMJ, 2018, 360: j5910. DOI: 10.1136/bmj.j5910.

[2] Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential [J]. Health Inf Sci Syst, 2014, 2: 3. DOI: 10.1186/2047-2501-2-3.

[3] Johnson AEW, Ppllard TJ, Shen L, et al. MIMIC-Ⅲ, a freely accessible critical care database [J]. Sci Data, 2016, 3: 160035. DOI: 10.1038/sdata.2016.35.

[4] Desautels T, Calvert J, Hoffman J, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach [J]. JMIR Med Inform, 2016, 4 (3): e28. DOI: 10.2196/medinform.5909.

[5] Calvert J, Mao Q, Hoffman JL, et al. Using electronic health record collected clinical variables to predict medical intensive care unit mortality [J]. Ann Med Surg (Lond), 2016, 11: 52-57. DOI: 10.1016/j.amsu.2016.09.002.

[6] Celi LA, Mark RG, Stone DJ, et al. "Big data" in the intensive care unit: closing the data loop [J]. Am J Respir Crit Care Med, 2013, 187 (11): 1157-1160. DOI: 10.1164/rccm.201212-2311ED.

[7] Lee J, Ribey E, Wallace JR. A web-based data visualization tool for the MIMIC-Ⅱ database [J]. BMC Med Inform Decis Mak, 2016, 16: 15. DOI: 10.1186/s12911-016-0256-9.

[8] Anon. Ninth revision of the International Classification of Diseases (ICD-9) [EB/OL]. [2017-12-21].

[9] Angus DC, Linde-Zwirble WT, Lidicker J, et al. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care [J]. Crit Care Med, 2001, 29 (7): 1303-1310. DOI: 10.1097/00003246-200107000-00002.

[10] R Core Team. R: a language and environment for statistical computing [CP/OL]. Vienna: R Foundation for Statistical Computing, 2017 (1997-07-23) [2017-04-21].

[11] 王才宏，刘纪宁，卢安阳，等. 3 183 例急危重症患者疾病谱分析 [J]. 川北医学院学报，2016, 31 (2): 235-236, 244. DOI: 10.3969/j.issn.1005-3697.2016.02.25.
Wang CH, Liu JN, Lu AY, et al. Analysis of diseases spectrum in 3 183 cases of critically ill patients [J]. J North Sichuan Med College, 2016, 31 (2): 235-236, 244. DOI: 10.3969/j.issn.1005-3697.2016.02.25.

[12] 张在其，陈文标，陈玮莹，等. 广州市 97 823 例院前急救患者流行病学分析 [J]. 中华危重病急救医学，2011, 23 (2): 99-103. DOI: 10.3760/cma.j.issn.1003-0603.2011.02.011.
Zhang ZQ, Chen WB, Chen WY, et al. The epidemiological characteristic of 97 823 cases of pre-hospital medical care in Guangzhou city [J]. Chin Crit Care Med, 2011, 23 (2): 99-103. DOI: 10.3760/cma.j.issn.1003-0603.2011.02.011.

[13] 黄树青，满达，巴特金，等. 呼和浩特市 2016 年院前急救患者疾病谱分布及流行病学特点：附 28 325 例病例报告 [J]. 中华危重病急救医学，2018, 30 (1): 78-82. DOI: 10.3760/cma.j.issn.2095-4352.2018.01.015.
Huang SQ, Man D, Ba TJ, et al. Distribution and epidemiological characteristics of disease spectrum in patients with pre-hospital care in Hohhot in 2016: a case analysis in 28 325 patients [J]. Chin Crit Care Med, 2018, 30 (1): 78-82. DOI: 10.3760/cma.j.issn.2095-4352.2018.01.015.

[14] 袁野，秦伟毅，卢勇，等. EICU 救治患者的疾病分类特点 [J]. 中国急救医学，2009, 29 (4): 356-357. DOI: 10.3969/j.issn.1002-1949.2009.04.022.
Yuan Y, Qin WY, Lu Y, et al. Patients treated by EICU with ICD [J]. Chin J Crit Care Med, 2009, 29 (4): 356-357. DOI: 10.3969/j.issn.1002-1949.2009.04.022.

[15] 张海涛，王芳. 医院重症监护室新生儿病例回顾性调查研究 [J]. 泰山医学院学报，2016, 37 (9): 1007-1009. DOI: 10.3969/j.issn.1004-7115.2016.09.013.
Zhang HT, Wang F. Epidemiologic investigation of hospitalized neonates at a neonatal intensive care unit [J]. J Taishan Med College, 2016, 37 (9): 1007-1009. DOI: 10.3969/j.issn.1004-7115.2016.09.013.

[16] Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers [J]. Stud Health Technol Inform, 2015, 216: 574-578. DOI: 10.3233/978-1-61499-564-7-574.

（收稿日期：2018-03-08）